

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:01:19 : Search time 31 Seconds
(without alignments)
272.514 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41

Sequence: 1 TPLGPKWPEVFGRLASPGF.....GEYANDQERRWTLTAPPGYR 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Database: 671580 seqs, 206047115 residues

Size: 16

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database:

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP_mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	175	4 Q9UMV3	Q9umv3 homo sapien
2	41	100.0	181	4 Q9ULC7	Q9ulc7 homo sapien
3	41	100.0	185	4 Q9UBP3	Q9ubp3 homo sapien
4	16	39.0	16	4 Q9UC48	Q9uc48 homo sapien

ALIGNMENTS

RESULT 1
Q9UMV3
AC Q9UMV3: PRELIMINARY: PRT: 175 AA.
DT 01-MAY-2000 (TREMUREL. 13, Created)
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)

or 102 (b) prior and
to oligo class

DE Mannose binding lectin-associated serine protease-2 related protein,
DE MAP19 (19kDa) precursor (Fragment).
GN MASP-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=99192764; PubMed=10092804;
RA Slover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
RA Jensenius J.C., Schwaebel W.J.;
RT "Two constituents of the initiation complex of the mannose-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene.";
RL J. Immunol. 162:3481-3490(1999).
CC -I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: Y18282; CAB50729.1; -
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.229; -
DR InterPro: IPR000152; ASX_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_Ca; 1.
DR PROSITE: PS00101; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_Ca; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
KW Repeat; Signal.
FT NON-TER 1 1
FT SIGNAL <1 5
FT CHAIN 6 175
FT FT
SQ SEQUENCE 175 AA: 19504 MW: 474FB56221CD28D1 CRC64;
Query Match 100.0%; Score 41; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 3.8e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 41
DB 6 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 46
RESULT 2
Q9ULC7
AC Q9ULC7: PRELIMINARY: PRT: 181 AA.
DT 01-MAY-2000 (TREMUREL. 13, Created)
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)
DE MBL-associated serine protease(MASP)-2 (Fragment).
GN MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Takahashi M., Fujita T.;
RT "Partial genomic structure of human MBL-associated serine protease
RT (MASP)-2 (from exon 1 to exon 5).";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AB033742; BAA85659.1; -
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.229; -

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF-2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat.
FT NON_TER 181
SQ SEQUENCE 181 AA; 20171 MW; 5FB229C21B89DDBC CRC64;

Query Match 100.0%; Score 41; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 3; 9e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 41
|||||
16 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 56

RESULT 3
Q9UBP3 PRELIMINARY; PRT; 185 AA.
AC Q9UBP3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannose binding lectin-associated serine protease-2 related protein,
MAP19 (19 kDa) precursor (19 kDa) (MBL-associated protein Map19).
GN MASP-2 OR SNAP OR MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=99192764; PubMed=10092804;
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
Jenssen J.C., Schwaible W.J.;
RT "Two constituents of the initiation complex of the mannose-binding
lectin activation pathway of complement are encoded by a single
structural gene."
RT Immunol. 162:3481-3490(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99262288; PubMed=10330290;
RA Takahashi M., Endo Y., Fujita T., Matsushita M.;
RT "A truncated form of mannose-binding lectin-associated serine
protease (MASP)-2 expressed by alternative polyadenylation is a
component of the lectin complement pathway."
RT Int. Immunol. 11:859-863(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Takahashi M., Fujita T.;
RT "Partial genomic structure of human MBL-associated serine protease
(MASP)-2 (from exon 1 to exon 5)."
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Park D., Kim B., Baek K., Yoon J.;
RT "Structure of Human MASP-2 Gene."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: Y18284; CAB50731.1; -;
DR EMBL: AB008047; BAA78616.1; -;
DR EMBL: AB033742; BAA83658.1; -;
DR EMBL: Y18281; CAB50728.1; -;
DR EMBL: Y18283; CAB50730.1; -;

DR EMBL: Y18286; CAB50732.1; -;
DR EMBL: Y18287; CAB50734.1; -;
DR EMBL: AF321558; AAG50275.1; -;
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.229; -;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF-2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 16 185
FT
FT
FT
SQ SEQUENCE 185 AA; 20629 MW; 73B13D56FB229C2 CRC64;

Query Match 100.0%; Score 41; DB 4; Length 185;
Best Local Similarity 100.0%; Pred. No. 4e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 41
|||||
Db 16 TPLGPKWPEVFGRLASPGFGEYANDQERRWTLTAPPGYR 56

RESULT 4
Q9UC48 PRELIMINARY; PRT; 16 AA.
AC Q9UC48:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 18.5 kDa tumor marker protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=96212543; PubMed=8618346;
RA Rasmussen H.H., Ornloft T.F., Wolf H., Celis J.E.;
RT "Towards a comprehensive database of proteins from the urine of
J. Biol. 155:2113-2119(1996)."
RL J. Biol. 155:2113-2119(1996).
SQ SEQUENCE 16 AA; 1751 MW; 289540537A1B4400 CRC64;

Query Match 39.0%; Score 16; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 6; 7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LASPGFGEYANDQER 30
|||||
Db 1 LASPGFGEYANDQER 16

Search completed: January 11, 2003, 11:05:43
Job time : 31 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:05:49 ; Search time 11 Seconds
(without alignments)
72.313 Million cell updates/sec

Title: US-09-874-198-1

Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPPGR 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Size: 16
118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 500 summaries

Database: Published Applications-AA:*

1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	41	US-09-874-198-1	Sequence 1, Appli
2	41	100.0	41	US-09-874-238-1	Sequence 1, Appli
3	41	100.0	686	US-09-874-198-2	Sequence 2, Appli
4	41	100.0	686	US-09-874-238-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-874-198-1
; Sequence 1, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jenssens, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002002

; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-198-1

Query Match 100.0%; Score 41; DB 10; Length 41;
Best Local Similarity 100.0%; Pred. No. 3.6e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41
Db 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41

RESULT 2
US-09-874-238-1
; Sequence 1, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jenssens, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-238-1

Query Match 100.0%; Score 41; DB 10; Length 41;
Best Local Similarity 100.0%; Pred. No. 3.6e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41
Db 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGR 41

RESULT 3
US-09-874-198-2
; Sequence 2, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jenssens, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8

```

: SOFTWARE: FastSDQ for Windows Version 4.0
:
: SEQ ID NO 2
:
: LENGTH: 686
:
: TYPE: prt
:
: ORGANISM: Homo sapiens
:
US-03-874-198-2

```

Query Match	100.0%;	Score 41;	DB 10;	Length 686;
Best Local Similarity	100.0%;	Pred. No. 4.2e-34;		
Matches 41; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

OY 1 TPLGPKMEPEVGRGLASPGFEGEYANDQERRMTLTAA^{PGYR} 41
D6 16 TPLGPKMEPEVGRGLASPGFEGEYANDQERRMTLTAA^{PGYR} 56

RESULT 4
US-09-874-238-2

Query Match	100.0%;	Score 41;	DB 10;	Length 686;
Best Local Similarity	100.0%;	Pred. No. 4.2e-34;		
Matches 41; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Oy	I TPLGPKPEPVYGRLASPGFPEGAYANDQERRMTLTAPGYR	41
Dd	16 TPLGPKPEPVYGRLASPGFPGEYANQERRMRTLTAAPGYR	56

```
completed: January 11, 2003, 11:10:13
me : 11 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:03:54 : Search time 17 Seconds
(without alignments)
70.961 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEVANDQERRMTLTAPGVR 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

sd: 262574 seqs, 29422922 residues

lze: 16

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database: Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/CTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
atches found						

Search completed: January 11, 2003, 11:06:46
Job time: 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:54:24 ; Search time 10 seconds
(without alignments)
170.053 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41

Sequence: 1 TPLGPKWPEVGRLASPGF.....GEVANDQERWTLTPPGYR 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Size: 16

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	686	1 MAS2_HUMAN	000187 homo sapien

ALIGNMENTS

Result 1
MAS2_HUMAN STANDARD; PRT; 686 AA.
10187: 075754; 091270; 098280;
30-MAY-2000 (Rel. 39, last sequence update)
15-JUN-2002 (Rel. 41, last annotation update)
DE Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)
DE (Mannan-binding protein associated serine protease 2) (MASP-2)
DE (MBL-associated serine protease 2).
GN MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97242412; PubMed=9087411;
RA Thiel S., Jensen T.V., Stover C.M., Schwaebel W.J., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "A second serine protease associated with mannan-binding lectin that
RT activates complement.";
RL Nature 386:506-510(1997).
RN [2]
RP SEQUENCE FROM N.A.

RA Thiel S., Vorup-Jensen T., Stover C.M., Schwaebel W.J., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "Identification and characterization of a novel protein of the human
RT complement system, mannan-binding lectin-associated serine protease-2
RT (MASP-2).";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99192764; PubMed=10092804;
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
RA Jensenius J.C., Schwaebel W.J.;
RT "Two constituents of the initiation complex of the mannan-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene";
RL J. Immunol. 162:3481-3490(1999).
RN [4]
RP SEQUENCE FROM N.A.
RA Park D., Kim B., Baek K., Yoon J.;
RT "Structure of human MASP-2 gene";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRYPSIN PROTEASE THAT PRESUMABLY PLAYS AN IMPORTANT ROLE
CC IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER
CC ACTIVATION IT CLEAVES C4 GENERATING C4A AND C4B.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y09926; CAA71059.1; -;
DR EMBL: X98400; CAA67050.1; -;
DR EMBL: Y18287; CAB50735.1; -;
DR EMBL: Y18286; CAB50733.1; -;
DR EMBL: AF321562; AAG50274.1; -;
DR EMBL: AF321558; AAG50274.1; JOINED.
DR EMBL: AF321559; AAG50274.1; JOINED.
DR EMBL: AF321560; AAG50274.1; JOINED.
DR EMBL: AF321561; AAG50274.1; JOINED.
DR HSP: P00763; IDPO.
DR MEROPS: S01.229; -;
DR GeneW: HGNC:6902; MASP2.
DR MIM: 605102; -;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser-protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00431; CUB; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; TRYP-SPC; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Complement pathway; Serine protease; Protease;

```

KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 686 MANNAN-BINDING LECTIN SERINE PROTEASE 2.
FT CHAIN 16 444 MANNAN-BINDING LECTIN SERINE PROTEASE 2 A
FT CHAIN 445 686 CHAIN.
FT CHAIN MANNAN-BINDING LECTIN SERINE PROTEASE 2 B
FT DOMAIN 16 137 CHAIN.
FT DOMAIN 138 181 CUB 1.
FT DOMAIN 184 296 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 299 362 CUB 2.
FT DOMAIN 365 431 SUSHI 1.
FT DOMAIN 445 686 SUSHI 2.
FT ACT_SITE 483 486 SERINE PROTEASE.
FT ACT_SITE 532 532 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 633 633 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT MOD_RES 158 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 72 90 HYDROXYLATION (POTENTIAL).
FT DISULFID 142 156 POTENTIAL.
FT DISULFID 152 165 POTENTIAL.
FT DISULFID 167 180 POTENTIAL.
FT DISULFID 184 211 POTENTIAL.
FT DISULFID 241 259 POTENTIAL.
FT DISULFID 300 348 POTENTIAL.
FT DISULFID 328 361 POTENTIAL.
FT DISULFID 366 412 POTENTIAL.
FT DISULFID 396 430 POTENTIAL.
FT DISULFID 434 552 INTERCHAIN (POTENTIAL).
FT DISULFID 598 618 POTENTIAL.
FT DISULFID 629 660 POTENTIAL.
FT CONFLICT 361 362 MISSING (IN REF. 3).
FT CONFLICT 371 371 D -> Y (IN REF. 4).
FT CONFLICT 372 372 L -> LCS (IN REF. 3).
FT CONFLICT 442 442 G -> E (IN REF. 4).
FT CONFLICT 447 447 G -> E (IN REF. 4).
FT CONFLICT 461 462 MISSING (IN REF. 3).
FT CONFLICT 473 473 L -> LIL (IN REF. 3).
SQ SEQUENCE 686 AA; 75685 MW; 4E34DED159448A2A CRC64;

```

Query Match Best Local Similarity 100.0%; Score 41; DB 1; Length 686;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPGGR 41
    |||
Db 16 TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPGGR 56

```

Completed: January 11, 2003, 11:04:59
e : 10 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 11:02:39 ; Search time 20 Seconds

(without alignments)
197.076 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEYANDQERRWTLTAPPGYR 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

hed: 283224 seqs, 96134422 residues

Size: 16

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 500 summaries

Database:

PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	41	100.0	686	1 A59271	Ra-reactive factor

ALIGNMENTS

RESULT 1

A59271
Ra-reactive factor (EC 3.4.21.-) 2 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
C:Accession: A59271
R:Thiel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaible, W.J.; Laursen, S.B.; Poulsen, K.
Nature 386, 506-510, 1997
A:Title: A second serine protease associated with mannan-binding lectin that activates C
A:Reference number: A59271; MUID:97242412; PMID:9087411
A:Accession: A59271
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-686 <JEN>
A:Cross-references: GB:Y09926; NID:q4007626; PIDN:CAW1059.1; PID:q4007627
A:Experimental source: tissue liver
A:Note: submitted to GenBank, December 1996
A:Note: parts of this sequence, including the amino end of the mature protein, were dete
C:Genetics:
A:Gene: GDB:MASP2
A:Cross-references: GDB:6071500
A:Map position: 1p36.2-1p36.3

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serin
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-444,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>
F:19-134/Domain: C1r/C1s repeat homology <C1R1>
F:142-180/Domain: EGF homology <EGF>
F:184-293/Domain: C1r/C1s repeat homology <C1R2>
F:300-361/Domain: complement factor H repeat homology <FH1>
F:366-430/Domain: complement factor H repeat homology <FH2>
F:445-679/Domain: trypsin homology <TRY>
F:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,366-430,434-5
F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:444-445/Cleavage site: Arg-Ile (autolytic) #status predicted
F:483,532,633/Active site: His, Asp, Ser #status predicted

Query Match	100.0%	Score 41:	DB 1:	Length 686:
Best Local Similarity	100.0%	Pred. No. 6.7e-35:		
Matches 41:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
Oy	1	TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGYR 41		
Db	16	TPLGPKWPEPVFGRLASPGFGEYANDQERRWTLTAPPGYR 56		

Search completed: January 11, 2003, 11:06:16
Job time : 20 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:53:09 ; Search time 36 Seconds

(without alignments)
151.758 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 41
Sequence: 1 TPLGPKWPEPVFGRLASPGF.....GEVANDQERRWTLTAPPGYR 41

Scoring table: OLIGO
Gapop 60.0 / Capext 60.0

hed: 908470 seqs, 133250620 residues

Size: 16

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 500 summaries

Database:

A.Geneseq_101002:*

- 1: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	41	AAE14563	Human 20 kDa MASP-
2	41	100.0	671	AAE14565	Human mature MASP-
3	41	100.0	686	AAE14564	Human MASP-2 prote
4	41	100.0	686	AAE14568	Human MASP-2 prote
5	41	100.0	1034	ABG21134	Novel human diago

ALIGNMENTS

RESULT 1
ID AAE14563 standard: peptide: 41 AA.
XX
AC AAE14563:
XX
DT 17-MAY-2002 (first entry)
XX
DE Human 20 kDa MASP-2 fragment.
XX
KW Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW complement fixation; infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
OS Homo sapiens.
XX
PN WO200206460-A2.
XX
PD 24-JAN-2002.
XX
PF 13-JUL-2001: 2001WO-DK00499.
XX
PR 13-JUL-2000: 2000DK-0001089.
XX
PR 01-JUN-2001: 2001DK-0000870.
XX
PA (JENS/) JENSENIUS J C.
XX (THIE/) THIEL S.
XX
PI Jensenius JC, Thiel S;
XX
DR WPI: 2002-179791/23.
XX
PT Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX
PS Claim 39: Page 3: 76pp: English.
XX
CC The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human MASP-2 20 kDa fragment.
XX
SQ Sequence 41 AA:
XX
Query Match 100.0%; Score 41: DB 23: Length 41:
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 41: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TPLGPKWPEPVFGRLASPGFGEVANDQERRWTLTAPPGYR 41
DB 1 TPLGPKWPEPVFGRLASPGFGEVANDQERRWTLTAPPGYR 41
RESULT 2
AAE14565
ID AAE14565 standard: peptide: 671 AA.
XX
AC AAE14565:
XX
XX
DT 17-MAY-2002 (first entry)

XX		Human mature MASP-2 protein.
DE		
XX		
KW		Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KM		complement fixation; infection; microbe; retrovirus; HIV; abortion;
KX		human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Region	15..671
FT		/note="This region is specifically claimed in
FT		claim 45"
XX		
PN	WO200206460-A2.	
PD		
XX	24-JAN-2002.	
XX		
XX	13-JUL-2001; 2001WO-DK00499.	
XX		
PA	3-JUL-2000; 2000DK-0001089.	
PA	31-JUN-2001; 2001DK-0000870.	
XX	(JENS/) JENSENIUS J C.	
XX	(THIE/) THIEL S.	
PI	Jensenius JC, Thiel S;	
DR	WPI: 2002-179791/23.	
XX		
PT	Use of a polypeptide comprising amino acid sequence derived from	
PT	mannan-binding lectin associated serine protease-2 (MASP-2) for	
PT	producing pharmaceutical composition, to treat bacterial, fungal, viral	
XX	infections -	
PS	Claim 45; Page 73-76; 76pp; English.	
XX		
CC	The invention relates to use of a polypeptide derived from	
CC	mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for	
CC	producing a pharmaceutical composition. MASP-2 is a complement-	
CC	-fixing enzyme and involved in lectin pathway of complement activation.	
CC	The pharmaceutical composition comprising MASP-2 is useful for	
CC	treating infectious caused by microbes such as fungus, yeast,	
CC	retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic	
CC	bacteria which are resistant to at least one antibiotic medicament or	
CC	multiresistant. The polynucleotide encoding MASP-2 is useful for	
CC	treating patients deficient in MASP-2. The invention also discloses	
CC	MASP-2 assays which are useful for determination of MASP-2 activity or	
CC	levels in patients suffering from e.g. infections, inflammatory disorders	
CC	and spontaneous recurrent abortion. The pharmaceutical composition	
CC	comprising MASP-2 inhibitor is useful for treating inflammatory	
XX	disorders. The present sequence is human mature MASP-2 protein.	
SQ	Sequence	671 AA:
	Query Match	100.0%; Score 41; DB 23; Length 671;
	Best Local Similarity	100.0%; Pred. No. 1.2e-33;
	Matches	41; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 TPUGKWPPEPVFGRIASPGFEGEYANDQRRMTLTAPPGYR 41	
Dd		
	1 TPUGKWPPEPVFGRIASPGFEGEYANDQRRMTLTAPPGYR 41	
RESULT 3		
AELI4564		
ID	AAEI4564 standard; Protein: 686 AA.	
AC	AAEI4564;	
DT	17-MAY-2002 (first entry)	
XX		
XX	Human MASP-2 protein.	

XX	Homo sapiens.	Location/Qualifiers
XX		
XX	Key	1..15
FX	Peptide	/label= Signal_peptide
FT		16..686
FT	Protein	/label= Mature_MASP-2_protein
FT		16..556
FT	Region	/note= "MASP-2 20 kDa fragment"
FT		16..31
FT	Domain	/label= C1r/C1s-like_domain
FT		16..296
FT	Region	/note= "This region is specifically claimed in claim 46"
FT		30..444
FT	Region	/note= "This region is specifically claimed in claim 41. The region 30 to 296 is specifically claimed in claim 47"
FT		108..134
FT	Region	/note= "Fragment obtained by amino acid sequencing of peptides"
FT		135..146
FT	Domain	/label= EGF-like_domain
FT		138..296
FT	Region	/note= "This region is specifically claimed in claim 43"
FT		155
FT	Misc-difference	/note= "Encoded by CA"
FT		156
FT	Misc-difference	/note= "Encoded by C"
FT		183..199
FT	Domain	/label= C1r/C1s-like_domain
FT		293..307
FT	Domain	/label= CCP-1_domain
FT		/note= "Complement control protein domain"
FT		363..376
FT	Domain	/label= CCP-2_domain
FT		377..388
FT	Region	/note= "Fragment obtained by amino acid sequencing of peptides"
FT		410..417
FT	Region	/note= "Fragment obtained by amino acid sequencing of peptides"
FT		432..441
FT	Region	/note= "Linker"
FT		445..468
FT	Domain	/label= Serine_protease_domain
FT		483
FT	Active-site	532
FT	Active-site	633
XX		
PN	WO200206460-A2.	
XX		
PD	24-JAN-2002.	
XX		
PF	13-JUL-2001; 2001WO-DK00499.	
XX		
PR	13-JUL-2000; 2000DK-0001089.	
PR	01-JUN-2001; 2001DK-0000870.	
XX		
PA	(JENS./) JENSENIUS J C.	
PA	(THIE./) THIEL S.	
XX		
XX	Jensenius JC, Thiel S;	
PI		
XX		
DR	WPI: 2002-179791/23.	
DR	N-PSDB: AAD24224.	
XX		
PT	Use of a polypeptide comprising amino acid sequence derived from	
PT	mannan-binding lectin associated serine protease-2 (MASP-2) for	
PT	producing pharmaceutical composition, to treat bacterial, fungal, viral	

PT Infections -
XX
PS Claim 41; Fig 6; 76pp; English.
XX
CC The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human MASP-2 protein.
CC Note: The present sequence is stated as being the same as
CC SEQ ID NO:2 shown in sequence listing of the specification (AAE14568).
XX However the sequences differ at various locations.
XX
SQ Sequence 686 AA:
XX
Query Match 100.0%; Score 41; DB 23; Length 686;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TPLGPKWPEPVFGRLASPCGFGEYANDQERRMTLTAPPGYR 41
16 TPLGPKWPEPVFGRLASPCGFGEYANDQERRMTLTAPPGYR 56
DB
XX
RESULT 4
AAE14568
ID AAE14568 standard; Protein; 686 AA.
XX
AC AAE14568;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human MASP-2 protein, alternative version.
XX
KW Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW complement fixation; infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..15
FT Protein 16..686
FT Protein 16..686
FT Misc-difference 155
FT Misc-difference 155 /label= Mature_MASP-2_protein
FT Misc-difference 156 /note= "Encoded by CA"
FT Misc-difference 156 /note= "Encoded by C"
FT Misc-difference 298 /note= "Encoded by CAG"
FT Misc-difference 299 /note= "Encoded by CAG"
FT Misc-difference 299 /note= "Encoded by CAG"
XX
XX WO200206460-A2.
XX
XX 24-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
XX
XX 01-JUN-2001; 2001DK-0000870.
XX

PA (JENS/) JENSENIUS J C.
PA (THIE/) THIEL S.
XX
XX
PI Jensenius JC, Thiel S;
XX
XX WPI: 2002-179791/23.
XX
XX N-PSDB; AAD24224.
XX
PT Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX
PS Claim 41; Page 71-73; 76pp; English.
XX
CC The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human MASP-2 protein.
CC Note: The present sequence is stated as being the same as
CC SEQ ID NO:2 shown in figure 6 of the specification (AAE14564).
XX However the sequences differ at various locations.
XX
SQ Sequence 686 AA:
XX
Query Match 100.0%; Score 41; DB 23; Length 686;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TPLGPKWPEPVFGRLASPCGFGEYANDQERRMTLTAPPGYR 41
16 TPLGPKWPEPVFGRLASPCGFGEYANDQERRMTLTAPPGYR 56
DB
XX
RESULT 5
ABG21134
ID ABG21134 standard; Protein; 1034 AA.
XX
AC ABG21134;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21125.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX

19	81	34.0	688	10	US-09-874-198-7	Se
----	----	------	-----	----	-----------------	----

Atts 124, 125
our Applications
14th No. 5th
no adequate
conting.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	238	100.0	41	10	US-09-874-198-1	Sequence 1, Appl
2	238	100.0	41	10	US-09-874-238-1	Sequence 1, Appl
3	238	100.0	686	10	US-09-874-198-2	Sequence 2, Appl
4	238	100.0	686	10	US-09-874-238-2	Sequence 2, Appl
5	93.5	39.3	174	10	US-09-764-853-563	Sequence 563, Appl
6	93	39.1	679	10	US-09-874-198-6	Sequence 6, Appl
7	93	39.1	679	10	US-09-874-238-6	Sequence 6, Appl
8	91.5	38.4	525	9	US-10-004-551-18	Sequence 18, Appl
9	91.5	38.4	525	9	US-10-004-551-20	Sequence 20, Appl
10	91.5	38.4	525	9	US-10-004-551-22	Sequence 22, Appl
11	91.5	38.4	526	9	US-10-004-551-24	Sequence 24, Appl
12	91.5	38.4	993	9	US-10-004-551-14	Sequence 14, Appl
13	91.5	38.4	994	9	US-10-004-551-16	Sequence 16, Appl
14	81.5	34.2	144	9	US-09-886-428-4	Sequence 4, Appl
15	81.5	34.2	144	9	US-10-041-406-5	Sequence 5, Appl
16	81	34.0	101	10	US-09-799-118-6	Sequence 6, Appl
17	81	34.0	351	10	US-09-925-302-611	Sequence 611, Appl
18	81	34.0	449	10	US-09-919-497-89	Sequence 89, Appl
19	81	34.0	688	10	US-09-874-198-7	Sequence 7, Appl

20	81	34.0	688	10	US-09-874-338-7	Sequence 7, Appl 1
21	81	34.0	705	9	US-09-808-602-94	Sequence 94, Appl 1
22	78	32.8	458	10	US-09-925-301-1282	Sequence 1282, Appl 1
23	78	32.8	556	10	US-09-729-674-92	Sequence 92, Appl 1
24	74.5	31.3	230	10	US-09-925-300-1416	Sequence 1416, Appl 1
25	74.5	31.3	760	10	US-09-925-301-1024	Sequence 1024, Appl 1
26	73.5	30.9	673	10	US-09-874-198-8	Sequence 8, Appl 1
27	73.5	30.9	673	10	US-09-874-238-8	Sequence 8, Appl 1
28	72	30.3	116	9	US-10-041-406-4	Sequence 4, Appl 1
29	70	29.4	997	10	US-09-747-371-3	Sequence 3, Appl 1
30	69	29.0	101	9	US-09-887-593-6	Sequence 6, Appl 1
31	69	29.0	113	9	US-09-852-209A-22	Sequence 22, Appl 1
32	69	29.0	113	9	US-10-086-623-12	Sequence 12, Appl 1
33	69	29.0	392	10	US-09-898-570-39	Sequence 39, Appl 1
34	69	29.0	730	10	US-09-850-048A-2	Sequence 2, Appl 1
35	69	29.0	986	10	US-09-285-385C-19	Sequence 19, Appl 1
36	69	29.0	999	10	US-09-747-371-2	Sequence 2, Appl 1
37	68	28.6	102	9	US-09-887-593-7	Sequence 7, Appl 1
38	68	28.6	986	10	US-09-850-048A-4	Sequence 4, Appl 1
39	68	28.6	1013	10	US-09-942-166-5	Sequence 5, Appl 1
40	68	28.6	1013	10	US-09-285-385C-20	Sequence 20, Appl 1
41	68	28.6	1013	10	US-09-285-385C-2	Sequence 2, Appl 1
42	67	28.2	1012	10	US-09-285-385C-4	Sequence 4, Appl 1
43	67	28.2	1013	10	US-09-942-166-3	Sequence 3, Appl 1
44	65.5	27.5	415	9	US-09-905-251A-104	Sequence 104, Appl 1
45	65.5	27.5	415	9	US-09-902-853-104	Sequence 104, Appl 1
46	65.5	27.5	415	9	US-09-907-824-104	Sequence 104, Appl 1
47	65.5	27.5	415	9	US-09-907-824-104	Sequence 104, Appl 1
48	65.5	27.5	415	9	US-09-904-011-104	Sequence 104, Appl 1
49	65.5	27.5	415	10	US-09-909-320-104	Sequence 104, Appl 1
50	65.5	27.5	415	10	US-09-909-088B-104	Sequence 104, Appl 1
51	63.5	26.7	459	10	US-09-789-561-97	Sequence 97, Appl 1
52	63	26.5	113	9	US-09-852-209A-23	Sequence 23, Appl 1
53	63	26.5	113	9	US-10-086-623-23	Sequence 23, Appl 1
54	63	26.5	923	12	US-10-104-440-2	Sequence 2, Appl 1
55	63	26.5	931	12	US-10-104-440-4	Sequence 4, Appl 1
56	60	25.2	77	12	US-10-067-122-13	Sequence 13, Appl 1
57	60	25.2	132	9	US-09-808-602-88	Sequence 88, Appl 1
58	60	25.2	270	9	US-09-732-242-2	Sequence 2, Appl 1
59	60	25.2	395	10	US-09-800-729-86	Sequence 86, Appl 1
60	60	25.2	452	9	US-09-808-602-21	Sequence 21, Appl 1

ALIGNMENTS

```

RESULT 1
US-09-874-198-1.
; Sequence 1, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, ANTIBODY
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-198-1
Query Match      100.0%; Score 238; DB 10; Length 41
Best Local Similarity 100.0%; Pred. No. 5,1e-22;
Matches 41; Conservative 0; Mismatches 0; Index 0;

```



```
FILE REFERENCE: 09011-002002
CURRENT APPLICATION NUMBER: US/09/874,198
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 679
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-198-6

Query Match
Best Local Similarity 45.2%; Score 93; DB 10; Length 679;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

11 VGRGLASPGFPGGEYANDQERRWTLTAPPGYR 41
      |||: |||: | | | | | | | | | |
8 MFGQIQSPGYPDPSYPSDSEVTWNITVPDGR 38

RESULT 7
US-09-874-238-6
Sequence 6, Application US/09874238
Patent No. US20020082209A1
GENERAL INFORMATION:
APPLICANT: Jensenius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
FILE REFERENCE: 09011-002003
CURRENT APPLICATION NUMBER: US/09/874,238
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 679
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-238-6

Query Match
Best Local Similarity 39.1%; Score 93; DB 10; Length 679;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

11 VGRGLASPGFPGGEYANDQERRWTLTAPPGYR 41
      |||: |||: | | | | | | | | | |
8 MFGQIQSPGYPDPSYPSDSEVTWNITVPDGR 38

RESULT 8
US-10-004-551-18
Sequence 18, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHIMKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
```

```
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: n 2077 can be A, G, C, or T.
US-10-004-551-18

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
Db 403 PFWDMSKEPVCIACGCGVIRNGTGRIVSPGPGNYNNLTCCHLLEAPPGQR 454

RESULT 9
US-10-004-551-20
Sequence 20, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHIMKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: n 2077 can be A, G, C, or T.
US-10-004-551-20

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
Db 403 PFWDMSKEPVCIACGCGVIRNGTGRIVSPGPGNYNNLTCCHLLEAPPGQR 454

RESULT 10
US-10-004-551-22
Sequence 22, Application US/10004551
Publication No. US20030004310A1
GENERAL INFORMATION:
APPLICANT: SHIMKETS, RICHARD A
APPLICANT: FERNANDES, ELMA
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 15966-559
CURRENT APPLICATION NUMBER: US/10/004,551
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 09/635,949
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
US-10-004-551-22

Query Match
Best Local Similarity 38.4%; Score 91.5; DB 9; Length 525;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

5 PKW--PEPV-----FGRLASPGFPGGEYANDQERRWTLTAPPGYR 41
      | | | | | | | | | | | | | | | | | | | | | |
```

Db 403 PWDSEKPEVCIAACGGVIRNGTTGRIVSPGPGNYSNNLTCHMLLEAPGQR 454

RESULT 11
US-10-004-551-24

; Sequence 24, Application US/10004551
; Publication No. US20030004310A1

; GENERAL INFORMATION:

; APPLICANT: SHIMKETS, RICHARD A

; APPLICANT: FERNANDES, ELMA

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

; FILE REFERENCE: 15366-559

; CURRENT APPLICATION NUMBER: US/10/004,551

; CURRENT FILING DATE: 2001-12-05

; PRIOR APPLICATION NUMBER: 09/635,949

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 24

; LENGTH: 526

; TYPE: PRT

; ORGANISM: Homo sapiens

; OTHER INFORMATION: n 2093 can be A, G, C, or T.

Query Match 38.4%; Score 91.5; DB 9; Length 526;

Best Local Similarity 40.4%; Pred. No. 0.0016;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

Qy 5 PKW--PEPV-----FGRLASPGFPGEXANDQERRMTLTAPPGYR 41

Db 403 PWDSEKPEVCIAACGGVIRNGTTGRIVSPGPGNYSNNLTCHMLLEAPGQR 454

RESULT 12
US-10-004-551-14

; Sequence 14, Application US/10004551

; Publication No. US20030004310A1

; GENERAL INFORMATION:

; APPLICANT: SHIMKETS, RICHARD A

; APPLICANT: FERNANDES, ELMA

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

; FILE REFERENCE: 15366-559

; CURRENT APPLICATION NUMBER: US/10/004,551

; CURRENT FILING DATE: 2001-12-05

; PRIOR APPLICATION NUMBER: 09/635,949

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 14

; LENGTH: 993

; TYPE: PRT

; ORGANISM: Homo sapiens

; OTHER INFORMATION: n 1755 can be A, G, C, or T

Query Match 38.4%; Score 91.5; DB 9; Length 993;

Best Local Similarity 40.4%; Pred. No. 0.0032;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

Qy 5 PKW--PEPV-----FGRLASPGFPGEXANDQERRMTLTAPPGYR 41

Db 403 PWDSEKPEVCIAACGGVIRNGTTGRIVSPGPGNYSNNLTCHMLLEAPGQR 454

RESULT 13
US-10-004-551-16

; Sequence 16, Application US/10004551

; Publication No. US20030004310A1

; GENERAL INFORMATION:

; APPLICANT: SHIMKETS, RICHARD A

; APPLICANT: FERNANDES, ELMA

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

; FILE REFERENCE: 15366-559

; CURRENT APPLICATION NUMBER: US/10/004,551

; CURRENT FILING DATE: 2001-12-05

; PRIOR APPLICATION NUMBER: 09/635,949

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 16

; LENGTH: 994

; TYPE: PRT

; ORGANISM: Homo sapiens

; OTHER INFORMATION: n 1755 can be A, G, C, or T.

Query Match 38.4%; Score 91.5; DB 9; Length 994;

Best Local Similarity 40.4%; Pred. No. 0.0032;

Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

Qy 5 PKW--PEPV-----FGRLASPGFPGEXANDQERRMTLTAPPGYR 41

Db 403 PWDSEKPEVCIAACGGVIRNGTTGRIVSPGPGNYSNNLTCHMLLEAPGQR 454

RESULT 14
US-09-886-429-4

; Sequence 4, Application US/09886429

; Patent No. US20020160371A1

; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana

; TITLE OF INVENTION: 56739, A NOVEL CUB DOMAIN CONTAINING

; FILE REFERENCE: 10448-072001

; CURRENT APPLICATION NUMBER: US/09/886,429

; CURRENT FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: 60/213,963

; PRIOR FILING DATE: 2000-06-23

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 144

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Consensus sequence

Query Match 34.2%; Score 81.5; DB 9; Length 144;

Best Local Similarity 50.0%; Pred. No. 0.006;

Matches 17; Conservative 5; Mismatches 7; Indels 5; Gaps 2;

Qy 13 GRLASPGFP---GE-YANDQERRMTLTAPPGYR 41

Db 16 GTTTSPTNPNSPSSGESYRNLECVMTISAPPGYR 49

RESULT 15
US-10-041-406-5

; Sequence 5, Application US/10041406

; Patent No. US20020164705A1

; GENERAL INFORMATION:

; APPLICANT: Bandaru, Rajasekhar

; TITLE OF INVENTION: 39362, A NOVEL CUB DOMAIN CONTAINING

; FILE REFERENCE: 10448-130001

; CURRENT APPLICATION NUMBER: US/10/041,406

; CURRENT FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: 60/260,286

; PRIOR FILING DATE: 2001-01-08

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 144

; TYPE: PRT


```

; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-874-198-7

```

```

Query Match          34.0%; Score 81; DB 10; Length 688;
Best Local Similarity 44.1%; Pred. No. 0.037;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 8 PEVYGRILASPGPEGEYANDQERRMTLTAPPGYR 41
   1 : || : || || || : || : || || ||
   5 POKLFGVTSPLFPKPYNNFETTTVTPTGYR 38

```

```

RESULT 20
US-09-874-238-7
; Sequence 7, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-874-238-7

```

```

Query Match          34.0%; Score 81; DB 10; Length 688;
Best Local Similarity 44.1%; Pred. No. 0.037;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
QY 8 PEVYGRILASPGPEGEYANDQERRMTLTAPPGYR 41
   1 : || : || || || || : || : || || ||
   5 POKLFGVTSPLFPKPYNNFETTTVTPTGYR 38

```

Search completed: January 11, 2003, 11:03:49
Job time : 61 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:46:19 ; Search time 66 Seconds
(without alignments)
18.278 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPISGKPEPVGRLSPGF.....GEVANDERRWTLPAPGYR 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	35.1	1290	1	US-08-470-350B-2
2	81	34.0	101	2	US-08-242-097-6
3	81	34.0	101	2	US-09-206-695-6
4	81	34.0	449	2	US-08-839-008-9
5	78	32.8	449	2	US-08-839-008-2
6	76	31.9	52	4	US-09-227-357-363
7	74	31.1	401	2	US-08-839-008-5
8	74	31.1	468	2	US-08-839-008-7
9	74	31.1	468	4	US-09-032-523-8
10	69	29.0	101	4	US-09-374-135-6
11	69	29.0	730	4	US-08-872-757-2
12	69	29.0	102	4	US-08-572-225-1
13	68	28.6	102	4	US-09-374-135-7
14	68	28.6	591	3	US-08-991-408-4
15	68	28.6	591	3	US-09-432-473-4
16	68	28.6	986	4	US-08-872-757-4
17	68	28.6	1013	2	US-08-866-650-5
18	68	28.6	1013	2	US-08-866-650-5
19	68	28.6	1013	3	US-08-921-287-5
20	68	28.6	1013	3	US-08-991-408-2
21	68	28.6	1013	4	US-09-240-473-5
22	67	28.2	1013	4	US-09-432-473-2
23	67	28.2	1013	4	US-08-866-650-3
24	67	28.2	1013	4	US-09-021-287-3
25	65.5	27.5	415	4	US-09-240-473-3
26	64	26.9	925	4	US-09-032-523-2
27	63	26.5	532	1	US-08-494-168-9

Not a sig w gland

ALIGNMENTS

28	63	26.5	909	3	US-08-936-135-18	Sequence 18, App1
29	63	26.5	922	4	US-09-116-473-4	Sequence 4, App1
30	63	26.5	923	3	US-08-936-135-6	Sequence 6, App1
31	63	26.5	926	3	US-08-936-135-20	Sequence 20, App1
32	61	25.6	906	3	US-08-936-135-22	Sequence 22, App1
33	61	25.6	906	3	US-08-936-135-24	Sequence 24, App1
34	61	25.6	909	3	US-08-936-135-8	Sequence 8, App1
35	61	25.6	909	3	US-08-936-135-10	Sequence 10, App1
36	61	25.6	914	3	US-08-936-135-12	Sequence 12, App1
37	61	25.6	926	3	US-08-936-135-14	Sequence 14, App1
38	61	25.6	931	3	US-08-936-135-16	Sequence 16, App1
39	60	25.2	1057	3	US-08-931-820-4	Sequence 4, App1
40	60	25.2	1078	3	US-08-963-825-21	Sequence 21, App1
41	60	25.2	1078	4	US-09-500-811-21	Sequence 21, App1
42	60	25.2	1078	4	US-09-500-811-21	Sequence 21, App1
43	60	25.2	1078	4	US-09-548-608-21	Sequence 21, App1
44	58	24.4	666	4	US-09-341-587-1	Sequence 1, App1
45	58	24.4	1785	4	US-09-341-587-3	Sequence 1, App1
46	57.5	24.2	277	2	US-08-024-868-2	Sequence 2, App1
47	57.5	24.2	277	2	US-08-242-097-2	Sequence 2, App1
48	57.5	24.2	277	4	US-09-206-695-2	Sequence 2, App1
49	57.5	24.2	277	4	US-09-000-179-1	Sequence 1, App1
50	57.5	24.2	277	5	PCT-US96-11995-1	Sequence 1, App1
51	57.5	24.2	546	1	US-08-494-168-10	Sequence 10, App1
52	54.5	22.9	446	2	US-08-922-171-3	Sequence 3, App1
53	54.5	22.9	472	2	US-08-922-171-2	Sequence 2, App1
54	54	22.7	39	1	US-08-534-342-5	Sequence 5, App1
55	54	22.7	39	1	US-08-534-342-6	Sequence 6, App1
56	54	22.7	39	1	US-08-675-140-5	Sequence 5, App1
57	54	22.7	39	1	US-08-675-140-6	Sequence 6, App1
58	53	22.3	250	4	US-09-134-001C-2869	Sequence 2869, App1
59	53	22.3	351	1	US-08-470-202-46	Sequence 46, App1
60	53	22.3	351	1	US-08-471-770-46	Sequence 46, App1

ALIGNMENTS

RESULT 1
US-08-470-350B-2
; Sequence 2, Application US/08470350B
; Patent No. 5684126
; GENERAL INFORMATION:
; APPLICANT: Li, Xiao
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08470,350B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wolfe, Susan A
; REGISTRATION NUMBER: 33,568
; REFERENCE/DOCKET NUMBER: 01107.48790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 1290 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-350B-2

Query Match 35.1%; Score 83.5; DB 1; Length 1290;
Best Local Similarity 31.4%; Pred. No. 0.022;
Matches 16; Conservative 5; Mismatches 17; Indels 13; Gaps 1;

QY 4 GPKWPEPVF-----GRLASPGPGEGYANDQERRMTLTAPPGYR 41
DB 825 GPWMTSPFVNYTCGGFLTGLSGQFSPPYPSGSPNNARCLMNIIEVPNNYR 875

RESULT 2
US-08-242-097-6
Sequence 6, Application US/08242097
Patent No. 5846763

GENERAL INFORMATION:
APPLICANT: Lee, Tae Ho
ATTORNEY/AGENT INFORMATION:
NAME: Wisniewski, Hans Georg
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
TITLE OF INVENTION: Therefore and Uses Thereof
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,097
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528

ORAMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 101 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-242-097-6

Query Match 34.0%; Score 81; DB 2; Length 101;
Best Local Similarity 44.1%; Pred. No. 0.0027;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 8 PEPVGRLASPGPGEGYANDQERRMTLTAPPGYR 41
DB 7 POKLFGFVTSPLFKPKYPNNFETTVITVPTGYR 40

RESULT 3
US-09-206-695-6
Sequence 6, Application US/09206695
Patent No. 6210905
GENERAL INFORMATION:
APPLICANT: Lee, Tae Ho
ATTORNEY/AGENT INFORMATION:
NAME: Wisniewski, Hans Georg

APPLICANT: Vilcek, Jan
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
TITLE OF INVENTION: Therefore and Uses Thereof
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/206,695
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/242,097
FILING DATE: 13-MAY-1994

ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 101 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-206-695-6

Query Match 34.0%; Score 81; DB 4; Length 101;
Best Local Similarity 44.1%; Pred. No. 0.0027;
Matches 15; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 8 PEPVGRLASPGPGEGYANDQERRMTLTAPPGYR 41
DB 7 POKLFGFVTSPLFKPKYPNNFETTVITVPTGYR 40

RESULT 4
US-08-839-008-9
Sequence 9, Application US/08839008
Patent No. 5916758

GENERAL INFORMATION:
APPLICANT: Hurdle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25


```

: TELEFAX: 415-854-7694
:
: TELE: 66141 PENNIE
:
: INFORMATION FOR SEQ ID NO:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 730 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
US-08-872-757-2

```

Query Match	29.0%;	Score 69;	DB 4;	Length 730;
Best Local Similarity	44.8%;	Pred. No. 0.8;		
Matches 13;	Conservative 4;	Mismatches 12;	Indels 0;	Gaps 0;

QY 13 GRASPGEYANDQERRWTLTAPPGYR 41
| : ||| : | : | | | |
Db 600 GSITSPGMPKEYPPNKNCIWQLVAPTQYR 628

Page 12
572-225-1
idence 1, Application US/085722225
tent No. 5807981

```

1  GENERAL INFORMATION:
2  - APPLICANT: Prockop, Darwin J.
3  - APPLICANT: Hojima, Yoshio
4  - APPLICANT: Li, Shi-wu
5  - APPLICANT: Sieron, Aleksander
6  - APPLICANT: Brenner, Mitch
7  - TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND ITS USE FOR
8  - TITLE OF INVENTION: DRUG DEVELOPMENT FOR THE TREATMENT OF DISEASE
9  - NUMBER OF SEQUENCES: 24
10 - CORRESPONDENCE ADDRESS:

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patentin Release #1.0, Version #1.330
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/572,225
8  FILING DATE: 13-DEC-1995

```

ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25, 227
REFERENCE/DOCKET NUMBER: 8389-0311
TELECOMMUNICATION INFORMATION:

```

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
;
; OS-08-572-225-1

```

Query Match	29.0%;	Score 69;	DB 1;	Length 788;
Best Local Similarity	44.8%;	Pred. No. 0.88;		
Matches 13; Conservative	4;	Mismatches 12;	Indels 0;	Gaps 0;

```

OY      13 GRLASPGFPGGEYANDQERRWTLTAPPGYR 41
          | : | | | | : | | | | |
Db      402 GSITSPGWPKKEYPPNKNCLWQLVAPTQYR 430

```

RESULT 13
US-09-374-135-7
; Sequence 7, Application US/09374135

GENERAL INFORMATION:
 APPLICANT: Afar, Daniel E.
 APPLICANT: Hubert, Rene S.
 APPLICANT: Leong, Kahan
 APPLICANT: Raitano, Arthur B.
 APPLICANT: Saffran, Douglas C.
 APPLICANT: [REDACTED]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

; SEQ ID NO 7
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Mouse
US-09-374-135-7

```

Query Match	28.68;	Score 68;	DB 4;	Length 102;
Best Local Similarity	34.58;	Pred. No. 0.12;		
Matches 10;	Conservative 9;	Mismatches 10;	Indels 0;	Gaps 0;

```
QY      13  GRLASPGFPGGEYANDQERRWLTTPAPGYR  41
      | : | | : | : | | : : | | : |
Db      1  GTITSPNWDKYPKSKECTWAISTPGHR  29
```

RESULT 14
US-08-991-408-4
; Sequence 4, Application US/08991408
; Patent No. 6008017
CURRENT INFORMATION

GENERAL INFORMATION:
APPLICANT: ARLETH, ANTHONY J.
APPLICANT: WILLETTTE, ROBERT N.
APPLICANT: ELISHOURBAGY, NABIL A.
APPLICANT: LI, XIAORONG
TITLE OF INVENTION: HUWAN CARDIAC/BRAIN TOLL-
LOID-LIKE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 4

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Diskette
3  COMPUTER: IBM Compatible
4  OPERATING SYSTEM: DOS
5  SOFTWARE: FASTEO for Windows Version 2.0
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/991,408

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/034,471
FILING DATE: 02-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50038
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-408-4

Query Match 28.6%; Score 68; DB 3; Length 591;
Best Local Similarity 37.9%; Pred. No. 0.85;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 13 GRLASPGFGEYANDQERRMTLTAPPGYR 41
DB 361 GLITSPNMPDKYPSRKECTWEISATPGHR 389

15
432-473-4
Sequence 4, Application US/09432473
Patent No. 6365715
GENERAL INFORMATION:
APPLICANT: ARLETH, ANTHONY J.
APPLICANT: WILLETT, ROBERT N.
APPLICANT: ELSHOURBAGY, MABIL A.
APPLICANT: LI, XIAOTONG
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE PROTEIN
FILE REFERENCE: ATG-50038-D1
CURRENT APPLICATION NUMBER: US/09/432,473
EARLIER FILING DATE: 1999-11-01
EARLIER APPLICATION NUMBER: 08/991,408
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 60/034,471
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 591
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-432-473-4

Query Match 28.6%; Score 68; DB 4; Length 591;
Best Local Similarity 37.9%; Pred. No. 0.85;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

13 GRLASPGFGEYANDQERRMTLTAPPGYR 41
DB 361 GLITSPNMPDKYPSRKECTWEISATPGHR 389

RESULT 16
US-08-872-757-4
Sequence 4, Application US/08872757
Patent No. 6258584
GENERAL INFORMATION:
APPLICANT: Prockop, Darwin J.
APPLICANT: Hojima, Yoshio
APPLICANT: LI, Shi-Wu
APPLICANT: Sieron, Alexander
TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
NUMBER OF INVENTION: PROCESSES, METHODS AND USES THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/872,757
FILING DATE: 10-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/609,187
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-028-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 986 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-872-757-4

Query Match 28.6%; Score 68; DB 4; Length 986;
Best Local Similarity 34.5%; Pred. No. 1.5;
Matches 10; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 13 GRLASPGFGEYANDQERRMTLTAPPGYR 41
DB 756 GTITSPNMPDKYPSRKECTWEISATPGHR 784

RESULT 17
US-08-866-650-5
Sequence 5, Application US/08866650
Patent No. 5939321
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tolloid-Like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Quarles & Brady
STREET: 1 South Plackney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/866,650
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296, 93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:44:28 ; Search time 66 Seconds
(without alignments)
59.720 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKMPEPVFGRLASPGF.....GEVANDQERRWTLTAPGGR 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ed: 283224 seqs, 96134422 residues

number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	686	1 A59271	Ra-reactive factor
2	93	39.1	699	1 I54763	hypothetical factor
3	85	35.7	610	2 T15761	hypothetical prote
4	83.5	35.1	1290	2 A51190	ebnerin precursor
5	81	34.0	449	2 A55362	procollagen I C-pr
6	81	34.0	705	1 C1HURB	complement subcomp
7	80.5	33.8	977	2 I52657	seizure-related pr
8	78	32.8	3623	2 T09456	intrinsic factor-B
9	74.5	31.3	688	1 C1HUS	complement subcomp
10	74	31.1	319	2 I51569	UVS-2 protein - A1
11	74	31.1	402	2 JH0403	procollagen I C-pr
12	72.5	30.5	695	1 S05008	complement subcomp
13	72	30.3	3623	2 T08618	intrinsic factor-B
14	71.5	30.0	1070	2 T31069	tolloid-BMP-1 like
15	69	29.0	730	1 BWHU1	procollagen C-endo
16	69	29.0	823	1 A58788	procollagen C-endo
17	69	29.0	986	1 B58788	procollagen C-endo
18	69	29.0	991	2 I49540	procollagen C-endo
19	69	29.0	1464	2 S58984	development protei
20	68	28.6	504	2 S56745	mucin (clone pGM31
21	68	28.6	767	2 T30018	hypothetical prote
22	68	28.6	1594	2 T30549	hensin - rabbit
23	68	28.6	2083	2 T42721	CRP-ductin-alpha p
24	67.5	28.4	694	2 JC6554	complement subcomp
25	67	28.2	288	2 T33224	hypothetical prote
26	67	28.2	707	2 JC3218	procollagen C-endo
27	65.5	27.5	77	2 S10114	hypothetical prote
28	65.5	27.5	171	2 A34493	collagen alpha 1(I
29	65.5	27.5	920	2 B34493	collagen alpha 1(I

Handwritten notes:
Hit No. 1 is
guessed to be
SPT Receptor
results.

30	65	27.3	1057	1	A39288	dorsal-ventral pat
31	27.3	3871	2	T22812	hypothetical prote	
32	63	26.5	343	2	T20614	hypothetical prote
33	63	26.5	391	2	T34284	hypothetical prote
34	63	26.5	1524	2	T30337	polypeptide - Afr1
35	63	26.5	1669	1	CGHU4B	conserved hypothet
36	62	26.1	980	2	E72035	CT590 hypothet
37	62	26.1	980	2	E86589	hypothetical prote
38	61	25.6	294	2	T29838	Xenopus -a - Africa
39	60.5	25.4	271	2	I51536	collagen alpha 1(I
40	60.5	25.4	886	2	I50694	intracellular PHB
41	60	25.2	418	2	G97360	intracellular PHB
42	60	25.2	418	2	A82578	collagen alpha 1(I
43	60	25.2	1466	1	CGHU7L	collagen alpha 1(I
44	59.5	25.0	1433	2	A46053	probable oxygenase
45	59	24.8	511	2	G95394	collagen alpha 2 c
46	59	24.8	636	2	S41067	hypothetical prote
47	59	24.8	1051	2	A35763	hypothetical prote
48	58.5	24.6	161	2	A83269	hypothetical prote
49	58	24.4	105	2	T18036	hypothetical prote
50	58	24.4	347	2	T20618	hypothetical prote
51	58	24.4	381	2	T20622	collagen alpha 1(I
52	58	24.4	1464	2	S59856	collagen alpha 1(I
53	58	24.4	1670	1	CGHU3B	collagen alpha 3(I
54	58	24.4	2403	2	A59386	sanko - human
55	57.5	24.2	223	2	T16654	hypothetical prote
56	57.5	24.2	275	2	JC6506	tumor necrosis fac
57	57.5	24.2	277	2	A41735	hyaluronate-bindin
58	57.5	24.2	349	2	T29931	hypothetical prote
59	57.5	24.2	1497	2	I49607	procollagen type V
60	57.5	24.2	1691	1	S22917	collagen alpha 5(I

ALIGNMENTS

RESULT 1
A59271
Ra-reactive factor (EC 3.4.21.-) 2 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence-Revision 19-May-2000 #text-change 16-Jun-2000
C:Accession: A59271
R:Thiel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaeble, W.J.; Laursen, S.B.; Poulsen, Nature 386, 506-510, 1997
A:Title: A second serine protease associated with mannan-binding lectin that activate
A:Reference number: A59271; M01D:97242412; PMID:9087411
A:Accession: A59271
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-686 <REN>
A:Cross-references: GB:Y09926; NID:94007626; PIDN:CAAF1059.1; PID:94007627
A:Experimental source: tissue liver
A:Note: submitted to GenBank, December 1996
A:Note: parts of this sequence, including the amino end of the mature protein, were d
C:Genetics:
A:Gene: GDB:MASP2
A:Cross-references: GDB:6071500
A:Map position: 1p36.2-1p36.3
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serin
F:1-15/Domain: signal sequence #status predicted <Stc>
F:16-444/Domain: 445-686/Product: Ra-reactive factor 2 #status predicted <MAT>
F:19-134/Domain: C1r/C1s repeat homology <C1R1>
F:142-180/Domain: EGF homology <EGF>
F:184-293/Domain: C1r/C1s repeat homology <C1R2>
F:300-361/Domain: complement factor H repeat homology <FH1>
F:366-430/Domain: complement factor H repeat homology <FH2>
F:445-679/Domain: trypsin homology <TRY>
F:72-90/142-156/152-165/167-180/184-211/241-259/300-348/328-361/366-412/396-430/434-5
F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:444-445/Cleavage site: Arg-Ile (autolytic) #status predicted
F:483,532,633/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 238; DB 1; Length 686;
Best Local Similarity 100.0%; Pred. No. 4.7e-21;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRSLASPGFPGFYANDQERRMTLTAPPGR 41
DB 16 TPLGPKWPEPVFGRSLASPGFPGFYANDQERRMTLTAPPGR 56

RESULT 2

154763

Ra-reactive factor (EC 3.4.21.-) 1 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000

C:Accession: I54763; JN0883
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
Int. Immunol. 6, 665-669, 1994
Title: Molecular characterization of a novel serine protease involved in activation of

Reference number: I54763; MUID:94289349; PMID:8018603
Accession: I54763

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-699 <SAR>
A:Cross-references: GB:D28593; NID:9790963; PIDN:BA005928.1; PID:9471128

R:Takada, F.; Takayama, Y.; Hattuse, H.; Kawakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993

A:Title: A new member of the C1s family of complement proteins found in a bactericidal

A:Reference number: JN0883; MUID:94059062; PMID:8240317

A:Accession: JN0883

A:Molecule type: mRNA

A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>

A:Cross-references: DBDJ:017225; NID:9439712; PIDN:BA004477.1; PID:9439713

A:Experimental source: liver

C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo

C:Genetics:

A:Gene: GDB:MASP1; GDB:CRARF; CRARF1; PRSS5; MASP

A:Cross-references: GDB:361104; GDB:330954; OMIM:600521

A:Map position: 3q27-3q28

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H

C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydro

F:1-11/Domain: Signal sequence #status predicted <SIS>

F:18-448/449-699/Product: Ra-reactive factor #status predicted <MANT>

F:19-135/Domain: C1r/C1s repeat homology <C1R1>

F:143-181/Domain: EGF homology <EGF>

F:185-294/Domain: C1r/C1s repeat homology <C1R2>

F:201-432/Domain: complement factor H repeat homology <RH1>

F:432/Domain: complement factor H repeat homology <RH2>

F:691/Domain: trypsin homology <TRY>

F:178/407/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted

F:490/552/646/Active site: His, Asp, Ser #status predicted

Query Match 39.1%; Score 93; DB 1; Length 699;
Best Local Similarity 45.2%; Pred. No. 0.0019;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 11 VGRSLASPGFPGFYANDQERRMTLTAPPGR 41
DB 27 MFGQIDSPGYPSPYSDSEVTNITVPDGR 57

RESULT 3

T16761

hypothetical protein R151.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001

C:Accession: T16761
R:Fulton, L.

submitted to the EMBL Data Library, February 1994

A:Description: The sequence of C. elegans cosmid R151.

A:Reference number: S44639

A:Accession: T16761

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-610 <FUI>

A:Cross-references: EMBL:000036; NID:9458996; PID:9459001; PIDN:AAA50653.1; CESP:R151

C:Genetics:

A:Gene: CESP:R151.5

A:Introns: 68/3; 120/1; 151/1; 307/1; 448/2; 525/2; 567/2; 592/2

Query Match 35.7%; Score 85; DB 2; Length 610;
Best Local Similarity 51.9%; Pred. No. 0.015;
Matches 14; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 14 RLASPGFPGFYANDQERRMTLTAPPGR 40
DB 383 RLSPGYPREFEGQCSWMLVAPGH 409

RESULT 4

A57190

Ebnerin precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999

C:Accession: A57190

R:Li, X.J.; Snyder, S.H.

J. Biol. Chem. 270, 17674-17679, 1995

A:Title: Molecular cloning of Ebnerin, a von Ebner's gland protein associated with ta

A:Reference number: A57190; MUID:9535352; PMID:7629065

A:Accession: A57190

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1290 <LTA>

A:Cross-references: GB:U32681; NID:97975346; PIDN:AA552248.1; PID:9975347

C:Superfamily: scavenger receptor cysteine-rich domain homology; C1r/C1s repeat homol

C:Keywords: extracellular protein; transmembrane protein

F:1-17/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC

F:142-245/Domain: scavenger receptor cysteine-rich domain homology <SRC1>

F:289-397/Domain: scavenger receptor cysteine-rich domain homology <SRC2>

F:583-688/Domain: C1r/C1s repeat homology <C1R1>

F:713-816/Domain: scavenger receptor cysteine-rich domain homology <SRC3>

F:838-944/Domain: C1r/C1s repeat homology <C1R2>

Query Match 35.1%; Score 83.5; DB 2; Length 1290;
Best Local Similarity 31.4%; Pred. No. 0.053; 17; Indels 13; Gaps 1;
Matches 16; Conservative 5; Mismatches 17; Indels 13; Gaps 1;

QY 4 GPKWPEPVF-----GRSLASPGFPGFYANDQERRMTLTAPPGR 41
DB 825 GPMVNSPPVNTVTCGFLGLSGQSSPYVPSYPNNAACLNIEVPNNYR 875

RESULT 5

A55362

procollagen I C-proteinase enhancer protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 05-Nov-1999

C:Accession: A55362

R:Takahara, K.; Kessler, E.; Blinamnov, L.; Brusel, M.; Eddy, R.L.; Jani-Sait, S.; S

J. Biol. Chem. 269, 26280-26285, 1994

A:Title: Type I procollagen COOH-terminal proteinase enhancer protein: identification

A:Reference number: A55362; MUID:95014462; PMID:7523404

A:Accession: A55362

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-449 <TAK>

A:Cross-references: GB:L33799; NID:9642907; PIDN:AAA61949.1; PID:9642908

C:Genetics:

A:Gene: GDB:PCOLCE

A:Cross-references: GDB:305468; OMIM:600270

A:Map position: 7q21.3-7q22

C:Superfamily C1r/C1s repeat homology
C:Keywords: extracellular protein; glycoprotein; proglutamic acid
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-443/Product: #status predicted <MAT>
F:37-146/Domain: C1r/C1s repeat homology <C1R1>
F:159-270/Domain: C1r/C1s repeat homology <C1R2>
F:26/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:29,431/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 34.0% Score 81 DB 2 Length 449;
Best Local Similarity 37.8%; Pred. No. 0.034;
Matches 17; Conservative 5; Mismatches 13; Indels 10; Caps 1;

Oy 5 PKWEPPVF-----GRLASPGFGEYANDDERRTLTAPPG 39
Db 28 PNYTRPVFLCGDVKGESGVVASEGFPNSYPNKECIWTITVPEG 72
| : ||| | : ||||| : ||||| : ||||| :
T 6
B
A:Element subcomponent C 1RBAR.GIF (EC 3.4.21.41) precursor [validated] - human
Species: Homo sapiens (man)
C:Date: 15-Nov-1994 #sequence, revision 30-Jun-1991 #text change 03-Jun-2002
C:Accession: A24170 #reference, A29768; S02422; A00916; A37820; S68830
R:Leytus, S.P.; Kurachi, K.; Sakariassen, K.S.; Davie, E.W.
Biochemistry 25, 4855-4863, 1986
A>Title: Nucleotide sequence of the cDNA coding for human complement C1r.
A:Reference number: A24170; MUID:87026566; PMID:3021205
A:Accession: A24170
A:Molecule type: mRNA
A:Residues: 1-705 <LEVY>
A:Cross-references: GB:M14058; NID:g179643; PIDN:AAA51851.1; PID:g179644
R:Journet, A.; Tosi, M.
Biochem. J. 240, 783-787, 1986
A>Title: Cloning and sequencing of full-length cDNA encoding the precursor of human comp
A:Reference number: A29768; MUID:87156625; PMID:3030286
A:Accession: A29768
A:Molecule type: mRNA
A:Residues: 1-151, 'L', 153-705 <JDQ>
A:Cross-references: GB:X04701; NID:g29538; PIDN:CAA28407.1; PID:g29539
R:Arlaud, G.J.; Willis, A.C.; Gagnon, J.
Biochem. J. 241, 711-720, 1987
A>Title: Complete amino acid sequence of the A chain of human complement-classical-pathw
A:Reference number: A29769; MUID:87241248; PMID:3036070
A:Accession: A29769
A:Molecule type: protein
A:Residues: 18-166, 'X', 168-463 <ARL>
Note: 152-Leu was also found
and G.J.: van Dorsselaer, A.; Bell, A.; Mancini, M.; Aude, C.; Gagnon, J.
Lett. 222, 129-134, 1987
A>Title: Identification of erythro-beta-hydroxyasparagine in the EGF-like domain of huma
A:Reference number: S02422; MUID:88005128; PMID:2820791
A:Accession: S02422
A:Molecule type: protein
A:Residues: 152-186 <ARL3>
Note: 152-Leu was also found
R:Arlaud, G.J.; Gagnon, J.
Biochemistry 22, 1758-1764, 1983
A>Title: Complete amino acid sequence of the catalytic chain of human complement subcomp
A:Reference number: A00916; MUID:83204782; PMID:6303394
A:Accession: A00916
A:Molecule type: protein
A:Residues: 464-705 <AR2>
R:Thielens, N.M.; Aude, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.
J. Biol. Chem. 265, 14469-14475, 1990
A>Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated NH
A:Reference number: A37820; MUID:90354439; PMID:2387866
A:Accession: A37820
A:Molecule type: protein
A:Residues: 18-26, 'L', 153-160; 'X', 252-255 <THI>
R:Pelloux, S.; Thielens, N.M.; Hudry-Clergeon, G.; Pettilot, Y.; Filhol, O.; Arlaud, G.J.
FEBS Lett. 386, 15-20, 1996
A>Title: Identification of a cryptic protein kinase CK2 phosphorylation site in human CD

```

A:Reference number: S68830; MUID:96221263; PMID:8635554
A:Accession: S68830
A:Molecule type: protein
A:Residues: 133-137;187-211;610-613 <PEL>
A:Experimental source: plasma
C:Comment: C1r is a dimer of identical chains, each of which is activated by cleavage A chain, while fragment gamma remains disulfide bonded to the B chain to form C1r II. C:Comment: This protein is a serine protease that combines with C1q and C1s to form Cn, activate C2 and C4.
C:Genetics:
A:Gene: GDB:C1R
A:Cross-references: GDB:119729; OMIM:216950
A:Map position: 12p13-12p13
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto C:Keywords: acute phase; beta-hydroxyaspartate; calcium binding; complement pathway; F:1-17/Domain: signal sequence #status predicted <SIG>
F:1-17/Domain: C1r/C1s repeat homology <C1RI>
F:18-463/Product: complement C1r chain A #status experimental <NCH>
F:146-189/Domain: EGF homology <EGF>
F:193-302/Domain: C1r/C1s repeat homology <C1RI2>
F:297-463/Product: C1r gamma fragment #status experimental <GFR>
F:309-311/Domain: complement factor H repeat homology <FHII>
F:376-447/Domain: complement factor H repeat homology <FH2>
F:464-705/Product: complement C1r chain B #status experimental <BCBI>
F:464-697/Domain: trypsin homology <TRY>
F:71-89,146-165,161-174,176-189,193-220,250-268,309-358,338-371,376-429,406-447,451-5 F:129,221,514,581/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:167/Modified site: erythro-beta-hydroxyaspartate (Asn) #status experimental
F:206/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status experime F:502,557,654/Active site: His, Asp, Ser #status predicted

Query Match          34.0%   Score 81;   DB 1;   Length 705;
Best Local Similarity 44.1%;   Pred. No. 0.055;
Matches    15;   Conservative     5;   Mismatches      14;   Indels       0;   Gaps        0;

OY      8 PEPVGRGLASPGFGCEVANDQERRMTLTAPPGYR 41
         | : ||| : ||| : ||| : ||| : ||| : |||
Db       22 POKLEGVTSPLEKPKYPNNETTIVTPNGYR 55

RESULT 7
seizure-related protein SEZ-6 precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 31-Mar-2000
C:Accession: I52657
R:Shimizu-Nishikawa, K.; Kajiwara, K.; Kimura, M.; Katsuki, M.; Sugaya, E. Brain Res. Mol. Brain Res. 28, 201-210, 1995
A>Title: Cloning and expression of SEZ-6, a brain-specific and seizure-related cDNA.
A:Reference number: I52657; MUID:95240392; PMID:7723619
A:Accession: I52657
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-977 <RS>
A:Cross-references: GI:D29763; NID:g6933909; PID:BAA06167.1; PID:g6933910
C:Superfamily: complement factor H repeat homology
F:354-409/Domain: complement factor H repeat homology <FH01>
F:529-586/Domain: complement factor H repeat homology <FH02>
F:707-762/Domain: complement factor H repeat homology <FH03>
F:768-827/Domain: complement factor H repeat homology <FH04>
F:835-892/Domain: complement factor H repeat homology <FH05>

Query Match          33.8%   Score 80.5;   DB 2;   Length 977;
Best Local Similarity 36.5%;   Pred. No. 0.09;
Matches    19;   Conservative     3;   Mismatches      15;   Indels       0;   Gaps        2;

OY      5 PKW--PEPV-----FGRLASPGFPGEYANDQERRMTLTAPPGYR 41
         | | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       400 PFWDQGEFVCIAACGCVIARNATGTIGVSPGPNGNSNNLTGMHLLPESQR 451

RESULT 8
```

T09456
Intrinsic factor-B12 receptor Cubilin precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Aug-2002
C:Accession: T09456
R:Koziyaki, R.; Kristiansen, M.; Slihtaroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N.
Blood 91, 3593-3600, 1998
A:Title: The human intrinsic factor-vitamin B12 receptor, cubilin: Molecular characterization.
A:Reference number: Z16677; MVID:98241400; PMID:9572993
A:Accession: T09456
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3623 <KOZ>
A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AC082612.1; PID:93929529
C:Genetics:
A:Map position: 10p12
C:Superfamily: Intrinsic factor-B12 receptor cubilin; EGF homology
C:Keywords: receptor; vitamin B12 uptake
C:Domain: 3623/Product: intrinsic factor-B12 receptor #status predicted <SIG>
467/Domain: EGF homology <EGF>
Query Match 32.8%; Score 78; DB 2; Length 3623;
Best Local Similarity 51.7%; Pred. No. 0.74;
Matches 15; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
Y 12 FGRLASPGFGEYANDQERRMTLAPPGY 40
Db 3403 FGRLSPGMPDNDKDCVTTLTAPNH 3431
RESULT 9
C:INHUS
complement subcomponent C 1SBR.GIF (EC 3.4.21.42) precursor [validated] - human
N:Alternate names: C1 esterase precursor
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Jun-2002
C:Accession: A40496; A27381; S002732; S05634; A05140; A25396; A38407; B37820
R:Kusumoto, H.; Hirose, S.; Sailer, J.P.; Hagen, F.S.; Kurachi, K.
Proc. Natl. Acad. Sci. U.S.A. 85, 7307-7311, 1988
A:Title: Human genes for complement components C1r and C1s in a close tail-to-tail array
A:Reference number: A40496; MVID:89017187; PMID:2459702
A:Accession: A40496
A:Molecule type: mRNA
A:Residues: 1-688 <KUS>
A:Cross-references: GB:J04080; NID:q179645; PIDN:AA51852.1; PID:q179646
R:Stefan, M.; Duponchel, C.; Meo, T.; Jullier, C.
Immunology 26, 8516-8524, 1987
A:Title: Complete cDNA sequence of human complement C1s and close physical linkage of the
A:Reference number: A27381; MVID:88163522; PMID:2831944
A:Accession: A27381
A:Molecule type: mRNA
A:Residues: 1-688 <TOS>
A:Cross-references: GB:MI8767; NID:q179647; PIDN:AA51853.1; PID:q179648
R:MacKinnon, C.M.; Carter, P.E.; Smyth, S.J.; Dunbar, B.; Fothergill, J.E.
Eur. J. Biochem. 169, 547-553, 1987
A:Title: Molecular cloning of cDNA for human complement component C1s. The complete amir
A:Reference number: S00224; MVID:88082788; PMID:3500856
A:Accession: S00224
A:Molecule type: mRNA
A:Residues: 1-688 <MAC>
A:Cross-references: EMBL:X06596; NID:929542; PIDN:CAA29817.1; PID:9763110
A:Accession: S26732
A:Molecule type: protein
A:Residues: 16-36;66-116;170-236;246-262;265-280;282-284;287-308;315-363;384-394;421-435
R:Tosi, M.; Duponchel, C.; Meo, T.; Couture-Tosi, E.
J. Mol. Biol. 208, 709-714, 1989
A:Title: Complement genes C1r and C1s feature an intronless serine protease domain close
A:Reference number: S05634; MVID:90040704; PMID:2553984
A:Accession: S05634
A:Status: not compared with conceptual translation
A:Molecule type: DNA

A:Residues: 356-513, 'G', 514-688 <TOS>
R:Carter, P.E.; Dunbar, B.; Fothergill, J.E.
Biochem. J. 215, 565-571, 1983
A:Title: The serine proteinase chain of human complement component C1s. Cyanogen brom
A:Reference number: A05140; MVID:84104122; PMID:6362661
A:Accession: A05140
A:Molecule type: protein
A:Residues: 438-483, 'X', 485-500;503-534;542-558;561-572, 'A', 574-601;617-623;626-644, 6
R:Syecher, S.E.; Nick, H.; Rickli, E.E.
Eur. J. Biochem. 156, 49-57, 1986
A:Title: Human complement component C1s. Partial sequence determination of the heavy
A:Reference number: A25396; MVID:86164350; PMID:3007145
A:Accession: A25396
A:Molecule type: protein
A:Residues: 16-61;168-219;287-293, 'K', 295-334;384-445 <SPY>
R:Hess, D.; Schaller, J.; Rickli, E.E.
Biochemistry 30, 2827-2833, 1991
A:Title: Identification of the disulfide bonds of human complement C1s.
A:Reference number: A38407; MVID:91175725; PMID:2007122
A:Accession: A38407
A:Molecule type: protein
A:Residues: 131-134, 'X', 136-146, 'X', 148-150;155, 'X', 157-162;166-170, 'X', 172-174, 'X', 1
'X', 387-402, 'X', 404-408;416-424, 'X', 426-431;547-556;592-597;617, 'X', 619-627, 'X', 629-
R:Thielems, N.M.; Audé, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.
J. Biol. Chem. 265, 14469-14475, 1990
A:Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated
A:Reference number: A37820; MVID:90354439; PMID:2387866
A:Accession: B37820
A:Molecule type: protein
A:Residues: 16-25, 'X', 203-207 <THI>
R:Thielems, N.M.; Van Dorsselaer, A.; Gagnon, J.; Arlaud, G.J.
Biochemistry 29, 3570-3578, 1990
A:Title: Chemical and functional characterization of a fragment of C1s containing the
A:Reference number: A32672; MVID:90283368; PMID:2141278
A:Contents: annotation; erythro-beta-hydroxyasparagine site, content
A:Note: about half of the A chains contain erythro-beta-hydroxyasparagine
C:Comment: This protein is a serine proteinase that combines with C1q and C1r to form
C2 and C4.
C:Comment: C1s is a dimer of identical chains, each of which is activated by cleavage
C:Genetics:
A:Gene: GDB:C1S
A:Cross-references: GDB:119730; OMIM:120580
A:Map position: 12p13-12p13
A:Introns: 291/1; 329/3; 356/1; 399/1; 424/1
A:Note: the list of introns may be incomplete
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C:Keywords: acute phase; beta-hydroxyasparagine; calcium binding; complement pathway;
F:1-15/Domain: signal sequence #status predicted <SIG>
F:11-127/Domain: C1r/C1s repeat homology <C1R1>
F:16-688/Product: complement subcomponent C1s #status experimental <MAT>
F:16-437/Product: complement subcomponent C1s chain A (heavy chain) #status experimen
F:135-171/Domain: EGF homology <EGF>
F:175-287/Domain: C1r/C1s repeat homology <C1R2>
F:294-354/Domain: complement factor H repeat homology <FH1>
F:359-421/Domain: complement factor H repeat homology <FH2>
F:438-688/Product: complement subcomponent C1s chain B (light chain) #status experimen
F:438-675/Domain: trypsin homology <TRY>
F:65-83;135-147;143-156;158-171;175-202;234-251;294-341;321-354;359-403;386-421;425-5
F:149/Modified site: erythro-beta-hydroxyasparagine (Asn) (partial) #status experimen
F:174;406/Binding site: carbonyl (Asn) (covalent) #status experimental
F:437-438/Cleavage site: Arg-116 (complement subcomponent C1r) #status experimental
F:475;529;632/Active site: His, Asp, Ser #status predicted
Query Match 31.3%; Score 74.5; DB 1; Length 688;
Best Local Similarity 34.3%; Pred. No. 0.33; Mismatches 13; Indels 1; Gaps 1;
Matches 12; Conservative 9;
Y 7 WPEFVGRASPGEYANDQERRMTLAPPGY 40
Db 14 YAEPTMYGELTSPYPOAYSEVEKSMIDIEVPEGY 48
RESULT 10

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1070 <LID>
 A:Cross-references: EMBL:U57369; NID:g1899041; PID:g1899042; PIDN:AA047485.1
 C:Superfamily: dorsal-ventral patterning protein tolloid; astacin homology: C1r/C1s rep

Query Match 30.0%; Score 71.5; DB 2; Length 1070;
 Best Local Similarity 39.5%; Pred. No. 1.2;
 Matches 15; Conservative 5; Mismatches 17; Indels 1; Gaps 1;

OY 4 GPKPEVFGRLSPGPEYANDQERRMTLTAPPGR 41
 Db 521 GGNNPGE-GFLNSPAYDEYSGDKCEWITVREGYQ 557

RESULT 15

BMH1
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, splice form BMP1 - human
 N:Alternate names: bone morphogenic protein 1 (BMP1)

A:Species: Homo sapiens (man)
 A:Accession: A37278; E58788
 A:Title: 16-Sep-1992 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999

A:Residues: 1-730 <WOZ>
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500
 C:Genetics:

A:Gene: GDB:BMP1
 A:Cross-references: GDB:125203; OMIM:112264
 A:Map position: 8p21-8p21

C:Function:
 A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen type
 C:Superfamily: procollagen C-endopeptidase; astacin homology: C1r/C1s repeat homology; B

C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; g1
 F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-730/Product: procollagen C-endopeptidase splice form BMP1 #status predicted <MAT>

F:130-321/Domain: astacin homology <AST>
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>

F:551-587/Domain: EGF homology <EGF>
 F:591-700/Domain: C1r/C1s repeat homology <C1R3>
 F:91-142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644-66
 F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted

A:Title: Novel regulators of bone formation: molecular clones and activities.
 A:Reference number: A37278; MUID:89072730; PMID:3201241

A:Map position: 8p21-8p21
 C:Function:
 A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t

C:Superfamily: procollagen C-endopeptidase; astacin homology: C1r/C1s repeat homology
 C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication;

F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-730/Product: procollagen C-endopeptidase splice form BMP1 #status predicted <MAT>

F:130-321/Domain: astacin homology <AST>
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>

F:551-587/Domain: EGF homology <EGF>
 F:591-700/Domain: C1r/C1s repeat homology <C1R3>
 F:91-142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644-66
 F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted

A:Title: Novel regulators of bone formation: molecular clones and activities.
 A:Reference number: A37278; MUID:89072730; PMID:3201241

A:Map position: 8p21-8p21
 C:Function:
 A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t

C:Superfamily: procollagen C-endopeptidase; astacin homology: C1r/C1s repeat homology
 C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication;

A:Residues: 1-702, 'EKRPALQPPGRPHQLKFRVQKRNTPQ' <WOZ>
 A:Cross-references: GB:M22488; NID:g179499; PIDN:AAA51833.1; PID:g179500
 R:Takahara, K.; Lyons, G.E.; Greenspan, D.S.

J. Biol. Chem. 269, 32572-32578, 1994
 A:Title: Bone morphogenetic protein-1 and a mammalian tolloid homologue (mTld) are en

A:Reference number: A58788; MUID:95096114; PMID:7798260
 A:Molecule type: mRNA
 A:Accession: A58788

A:Cross-references: GB:U35278; NID:g619423; PIDN:AA041703.1; PID:g619424
 C:Genetics:

A:Gene: GDB:BMP1; BMP-1
 A:Cross-references: GDB:125203; OMIM:112264
 A:Map position: 8p21-8p21

C:Function:
 A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t

C:Superfamily: procollagen C-endopeptidase; astacin homology: C1r/C1s repeat homology
 C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication;

F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-823/Product: procollagen C-endopeptidase splice form HIS #status predicted <MAT>

F:130-321/Domain: astacin homology <AST>
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>

F:551-587/Domain: EGF homology <EGF>
 F:591-700/Domain: C1r/C1s repeat homology <C1R3>
 F:738-752/Region: histidine-rich

F:91,142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644

F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted
 F:214/Active site: Glu #status predicted
 F:565/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 29.0%; Score 69; DB 1; Length 823;
 Best Local Similarity 44.8%; Pred. No. 1.9;
 Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 13 GRASPFGPEYANDQERRMTLTAPPGR 41
 Db 600 GSITSPGMPREYPPNKNKCIQWLVAPQYR 628

RESULT 17
 B58788
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, tolloid-like splice form - huma

N:Alternate names: bone morphogenic protein 1, tolloid-like splice form
 C:Species: Homo sapiens (man)
 C:Date: 28-Mar-1998 #sequence_revision 09-Apr-1998 #text_change 16-Jul-1999

C:Accession: A37278; B58788
 R:Mooney, J.M.; Rosen, V.; Celeste, A.J.; Mltsock, L.M.; Whitters, M.J.; Kriz, R.W.;

A:Title: Novel regulators of bone formation: molecular clones and activities.
 A:Reference number: A37278; MUID:89072730; PMID:3201241

A:Map position: 8p21-8p21
 C:Function:
 A:Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen t

C:Superfamily: procollagen C-endopeptidase; astacin homology: C1r/C1s repeat homology
 C:Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication;

F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-730/Product: procollagen C-endopeptidase splice form BMP1 #status predicted <MAT>

F:130-321/Domain: astacin homology <AST>
 F:322-431/Domain: C1r/C1s repeat homology <C1R1>
 F:435-544/Domain: C1r/C1s repeat homology <C1R2>

F:551-587/Domain: EGF homology <EGF>
 F:591-700/Domain: C1r/C1s repeat homology <C1R3>
 F:91-142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644-66
 F:213,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted

F:23-866/Product: procollagen C-endopeptidase tolloid-like splice form #status predicted
F:130-321/Domain: astacin homology <AST>
F:322-431/Domain: Clr/Cls repeat homology <CLR1>
F:435-544/Domain: Clr/Cls repeat homology <CLR2>
F:551-587/Domain: EGF homology <EG1>
F:591-700/Domain: Clr/Cls repeat homology <CLR3>
F:707-742/Domain: EGF homology <EG2>
F:747-856/Domain: Clr/Cls repeat homology <CLR4>
F:860-973/Domain: Clr/Cls repeat homology <CLR5>
F:91-142,332,363,599/Binding site: carbohydrate (asn) (covalent) #status predicted
F:163-319,185-205,322-348,375-397,435-461,488-510,551-563,559-572,574-587,591-617,644-66
F:113,217,223,272/Binding site: zinc (His, His, Tyr) #status predicted
F:214/Active site: Glu #status predicted
F:565,720/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 29.0% Score 69; DB 1; Length 986;
Best Local Similarity 44.8% Pred. No. 2.3;
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

13 GRLASPGFPGEVANDQERRWTLTAPPGYR 41
I : I I I I I : : I I I I I
600 GSITSPGMPKREYPPNNKNCIMQLVAPPTQYR 628

RESULT 18
149540
procollagen C-endopeptidase (EC 3.4.24.19) precursor, tolloid-like splice form - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jan-2000
C:Accession: I49540
R:Fukagawa, M.; Noboru, S.; Hogan, B.L.M.; Jones, C.M.
Dev. Biol. 163, 175-183, 1994
A:Title: Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1) which is rel
A:Reference number: I49540; MUID:94229342; PMID:8174772
A:Accession: I49540
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:I24755; NID:9439606; PIDN:AAA37306.1; PID:9439607
C:Genetics:
A:Gene: bmp-1
C:Superfamily: procollagen C-endopeptidase; astacin homology; Clr/Cls repeat homology; E
C:Keywords: hydrolase; metalloproteinase; zinc
F:135-326/Domain: astacin homology <AST>
F:556-592/Domain: EGF homology <EG1>
F:596-705/Domain: Clr/Cls repeat homology <CLR>
F:812-747/Domain: EGF homology <EG2>
F:12-222,228,277/Binding site: zinc (His, His, His, Tyr) #status predicted
Active site: Glu #status predicted

Query Match 29.0% Score 69; DB 2; Length 991;
Best Local Similarity 44.8% Pred. No. 2.3;
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 13 GRLASPGFPGEVANDQERRWTLTAPPGYR 41
I : I I I I I : : I I I I I
Db 605 GSITSPGMPKREYPPNNKNCIMQLVAPPTQYR 633

RESULT 19
S58984
development protein tolkin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: S58984
R:Finelli, A.L.; Xie, T.; Bossie, C.A.; Blackman, R.K.; Padgett, R.W.
Genetics 141, 271-281, 1995
A:Title: The tolkin gene is a tolloid/BMP-1 homologue that is essential for Drosophila d
A:Reference number: S58984; MUID:96042912; PMID:8536976
A:Accession: S58984
A:Molecule type: mRNA
A:Residues: 1-1464 <FIN>
A:Cross-references: EMBL:U34777; NID:g1002985; PIDN:AAC47015.1; PID:g1002986

A:Note: the authors did not translate the codon for residue 722
C:Genetics:
A:Gene: tolkin
A:Cross-references: FlyBase:Fgn0004885
C:Superfamily: astacin homology; EGF homology
C:Keywords: hydrolase; metalloproteinase; zinc
F:529-722/Domain: astacin homology <AST>
F:958-993/Domain: EGF homology <EGF>
F:118-1153/Domain: EGF homology <EGF1>
F:614,618,624,673/Binding site: zinc (His, His, His, Tyr) #status predicted
F:615/Active site: Glu #status predicted

Query Match 29.0% Score 69; DB 2; Length 1464;
Best Local Similarity 41.4% Pred. No. 3.5;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

OY 13 GRLASPGFPGEVANDQERRWTLTAPPGYR 41
I I I I I : : I I I I I
Db 851 GRLESPNYPLDYLPNKECVWKITVPESTYQ 879

RESULT 20
S56745
mucin (clone pgm31-1) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Apr-2000
C:Accession: S56745
R:Nunes, D.P.; Keates, A.C.; Afshar, N.H.; Offner, G.D.
Biochem. J. 310, 41-48, 1995
A:Title: Bovine gall-bladder mucin contains two distinct tandem repeating sequences:
A:Reference number: S56744; MUID:95374471; PMID:7646470
A:Accession: S56745
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-504 <MUN>
A:Cross-references: GB:S78869; NID:g1042038; PIDN:AA835070.1; PID:g1042039
C:Superfamily: Clr/Cls repeat homology; scavenger receptor cysteine-rich domain homol
F:197-305/Domain: Clr/Cls repeat homology <CLR2>
F:311-417/Domain: scavenger receptor cysteine-rich domain homology <SRC>

Query Match 28.6% Score 68; DB 2; Length 504;
Best Local Similarity 39.3% Pred. No. 1.5;
Matches 11; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

OY 13 GRLASPGFPGEVANDQERRWTLTAPPGYR 40
I I I I I : : I I I I I
Db 206 GNFCSPSPGYPPNNAQCVWEIOVNPY 233

Search completed: January 11, 2003, 10:53:01
Job time : 69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:42:59 : Search time 121 Seconds
(without alignments)
69,818 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238
Sequence: 1 TPLGPKWPEPVGRLASPGF.....GEVANDERRWTLTAPGVR 41

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

ed: 671580 seqs, 206047115 residues

number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 60 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	175	4 Q9UMV3	Q9umv3 homo sapien
2	238	100.0	181	4 Q9UDC7	Q9udc7 homo sapien
3	238	100.0	185	4 Q9UBP3	Q9ubp3 homo sapien
4	183	76.9	185	11 Q9QXD4	Q9qxd4 mus musculu
5	183	76.9	190	11 Q9QXD5	Q9qxd5 mus musculu
6	183	76.9	367	11 Q9QXD2	Q9qxd2 mus musculu
7	183	76.9	685	11 Q92338	Q92338 mus musculu
8	183	76.9	685	11 Q91WP0	Q91wp0 mus musculu
9	180	75.6	678	11 Q9JJS8	Q9jjs8 rattus norv
10	170	71.4	186	11 Q9QXD4	Q9qxd4 rattus norv
11	164	68.9	177	11 Q9QXD6	Q9qxd6 rattus norv
12	164	68.9	185	11 Q9QXD8	Q9qxd8 rattus norv
13	164	68.9	216	11 Q9QXD8	Q9qxd8 rattus norv
14	100	42.0	688	13 Q9PVY4	Q9pvy4 xenopus lae
15	93	39.1	212	11 Q9Q020	Q9q020 rattus norv
16	93	39.1	701	11 Q9JJS9	Q9jjs9 rattus norv

Hit No. 3 has
best biology
including in
notations of
band 2

ALIGNMENTS

17	93	39.1	728	4 Q96RS4	Q96rs4 homo sapien
18	93	39.1	733	11 Q920S0	Q920s0 mus musculu
19	91.5	38.4	996	4 Q8TD25	Q8td25 homo sapien
20	87	36.6	16	4 Q9UC48	Q9uc48 homo sapien
21	84	35.3	3138	5 Q9VTP0	Q9vtp0 drosophila
22	83.5	35.1	384	13 Q9W630	Q9w630 cyprinus ca
23	83.5	35.1	1290	11 Q62827	Q62827 rattus norv
24	83	34.9	719	13 Q9PVY2	Q9pvy2 tritakis scy
25	81.5	34.2	745	13 Q9PVY3	Q9pvy3 cyprinus ca
26	80.5	33.8	605	11 Q62224	Q62224 mus musculu
27	80.5	33.8	977	11 Q62269	Q62269 mus musculu
28	80.5	33.8	991	11 Q62223	Q62223 mus musculu
29	80	33.6	832	11 Q9JLB4	Q9jlb4 mus musculu
30	78	32.8	1048	4 Q96B27	Q96b27 homo sapien
31	78	32.8	3389	4 Q96B09	Q96b09 homo sapien
32	78	32.8	3508	4 Q96RM4	Q96rm4 homo sapien
33	78	32.8	3564	11 Q923L3	Q923l3 mus musculu
34	78	32.8	3623	4 Q60494	Q60494 homo sapien
35	78	32.8	4215	5 Q9W332	Q9w332 drosophila
36	76	31.9	722	13 Q9PS25	Q9ps25 lampetra ja
37	76	31.9	3620	6 Q9TU53	Q9tu53 canis famil
38	75	31.5	421	11 Q89002	Q89002 mus musculu
39	75	31.5	707	11 Q9ET60	Q9et60 mus musculu
40	75	31.5	707	11 Q99K16	Q99k16 mus sapien
41	74.5	31.3	139	4 Q96GN2	Q96gn2 homo sapien
42	73.5	30.9	273	5 Q97070	Q97070 ascaris suu
43	73.5	30.9	288	5 Q9U994	Q9u994 ascaris suu
44	73	30.7	107	4 Q9H4W1	Q9h4w1 homo sapien
45	73	30.7	170	4 Q9H4W2	Q9h4w2 homo sapien
46	72.5	30.5	686	13 Q9DGC2	Q9dgc2 cyprinus ca
47	72	30.3	3623	11 Q70244	Q70244 rattus norv
48	71.5	30.0	414	11 Q9CX06	Q9cx06 mus musculu
49	71.5	30.0	414	11 Q8R4W6	Q8r4w6 mus musculu
50	71.5	30.0	1070	5 P91972	P91972 aplysia cal
51	70.5	29.6	685	13 Q9DGC1	Q9dgc1 cyprinus ca
52	70.5	29.6	685	13 Q9DGC0	Q9dgc0 cyprinus ca
53	70	29.4	698	13 Q9PU71	Q9pu71 xenopus lae
54	70	29.4	997	11 Q9JUS0	Q9jus0 mus musculu
55	69	29.0	241	11 Q92135	Q92135 rattus norv
56	69	29.0	735	13 Q57381	Q57381 xenopus lae
57	69	29.0	999	4 Q9N036	Q9n036 homo sapien
58	69	29.0	1464	5 Q24132	Q24132 drosophila
59	69	29.0	1464	5 Q23995	Q23995 drosophila
60	69	29.0	1464	5 Q9VC47	Q9vc47 drosophila

RESULT 1

ID	Q9UMV3	PRELIMINARY:	PRT:	175 AA.
AC	Q9UMV3	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Mannose binding lectin-associated serine protease-2 related protein,			
DE	MAP1s (19kDa) precursor (Fragment).			
GN	MASP-2.			
OS	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI-TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RA	MEDLINE=9192764; PubMed=10092804;			
RA	Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,			
RA	Jensenius J.C., Schwaebler W.J.,			
RT	Two constituents of the initiation complex of the mannose-binding			
RT	lectin activation pathway of complement are encoded by a single			
RT	structural gene.";			
RL	J. Immunol. 162:3481-3490(1999).			

```

CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: Y18282; CAB50729.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.229; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease; Repeat.
KW Repeat: Signal.
FT SIGNAL 1 1
FT CHAIN 6 175
          POTENTIAL.
          MANNOSE BINDING LECTIN-ASSOCIATED SERINE
          PROTEASE-2 RELATED PROTEIN,, MAP19
          (19KDA).
SEQUENCE 175 AA; 19504 MW; 474FB5622ICD28D1 CRC64;

Query Match 100.0%; Score 238; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 3.1e-23;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPGLPKWPEPVFGRLASPGFGEVYANDDERMTLTAPGXR 41
   |||||||
DB 6 TPGLKWPPEPVFGRLASPGFGEVYANDDERMTLTAPGXR 46

RESULT 2
O9ULC7 PRELIMINARY: PRT: 181 AA.
AC O9ULC7:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MBL-associated serine protease(MASP) (Fragment).
GN MASP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Takahashi M., Fujita T.;
   Partial genomic structure of human MBL-associated serine protease
   (MASP)-2 (from exon 1 to exon 5).";
CC Submitted (OCF-1999) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AB033742; BAA85659.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.229; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat.
KW NON_TER 181
FT NON_TER 181
FT SEQUENCE 181 AA; 20171 MW; 5FB229C21B89DDBC CRC64;

Query Match 100.0%; Score 238; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 3.2e-23;

```

Qy	1	1	TPLGKWPEDVFGRLASGEGFGEYANDQERRWTLTAPPGYR	41
Db	16	16	TPLGKWPEDVFGRLASGEGFGEYANDQERRWTLTAPPGYR	56
RESULT 3				
Q9UBP3		PRELIMINARY;	PRT;	185 AA.
AC	Q9UBP3;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Mannose binding lectin-associated serine protease-2 related protein, MASP19 (19 kDa) precursor (19 kDa) (MBL-associated protein Masp19).			
GN	MAASP-2 OR SMAAP OR MASP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RX	MEDLINE=99192764; PubMed=10092804;			
RA	Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,			
RA	Jenssens J.C., Schaeble W.J.;			
RT	"Two constituents of the initiation complex of the mannose-binding			
RT	lectin activation pathway of complement are encoded by a single			
RT	structural gene.";			
RL	J. Immunol. 162:3481-3490(1999).			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99262288; PubMed=10330290;			
RA	Takahashi M., Endo Y., Fujita T., Matsushita M.;			
RT	"A truncated form of mannose-binding lectin-associated serine			
RT	protease (MASP)-2 expressed by alternative polyadenylation is a			
RT	component of the lectin complement pathway.";			
RL	Int. Immunol. 11:859-863(1999).			
RP	[3]			
RP	SEQUENCE FROM N.A.			
RA	Takahashi M., Fujita T.;			
RT	"Partial genomic structure of human MBL-associated serine protease			
RT	(MASP)-2 (from exon 1 to exon 5).";			
RL	Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Park D., Kim B., Baek K., Yoon J.;			
RT	"Structure of Human MASP-2 Gene.";			
RT	Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.			
CC	-I- SIMILARITY: CONTAINS 1 CUB DOMAIN.			
DR	EMBL: Y18284; CAB50731.1; -			
DR	EMBL: AB008047; BAA78616.1; -			
DR	EMBL: AB033742; BAA85658.1; -			
DR	EMBL: Y18281; CAB50728.1; -			
DR	EMBL: Y18283; CAB50730.1; -			
DR	EMBL: Y18286; CAB50732.1; -			
DR	EMBL: Y18287; CAB50734.1; -			
DR	EMBL: AF321558; AAG50275.1; -			
DR	HSSP: P00736; IAF0.			
DR	MEROPS: S01.229; -			
DR	InterPro: IPR000152; Asx_hydroxyl.			
DR	InterPro: IPR000859; CUB_domain.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR001881; EGF_CA.			
DR	Pfam: PF00431; CUB; 1.			
DR	SMART: SM00042; CUB; 1.			
DR	SMART: SM00179; EGF_CA; 1.			
DR	PROSITE: PS00010; ASX_HYDROXYL, UNKNOWN_1.			
DR	PROSITE: PS01180; CUB; 1.			
DR	PROSITE: PS01186; EGF_2; 1.			
DR	PROSITE: PS01187; EGF_CA; 1.			
KW	Calcium-binding; EGF-like domain; glycoprotein; lectin; Protease;			

KW	Repeat:Signal.			POTENTIAL.
FT	SIGNAL	1	15	MANNOSYL-BINDING LECTIN-ASSOCIATED SERINE
FT	CHAIN	16	185	PROTEASE-2 RELATED PROTEIN, MAP19
FT			(19 kDa).	
SO	SEQUENCE	185 AA;	20629 MW;	73B133D56FB229C2 CRC64;
	Query Match		100.0%;	Score 238; DB 4; Length 185;
	Best Local Similarity		100.0%;	Fid. No. 4.7e-16;
	Matches	41; Conservative	0; Mismatches	0; Indels
OY				Gaps
Dd	1 TPLGKWPPEVFGRLASPGFPEGYANDDERKWTLPAPGGR 41			
	16 TPLGKWPPEVFGRLASPGFPEGYANDDERKWTLPAPGGR 56			
RESULT 4				
090XA4	PRELIMINARY:		PRT: .	185 AA.
090XAD				
01-MAY-2000 (TRENBLREL. 13, Created)				
01-MAY-2000 (TRENBLREL. 13, Last sequence update)				
01-JUN-2002 (TRENBLREL. 21, Last annotation update)				
Mouse MAP19 precursor.				
MASP2 OR MASP-2/MAP19.				
Mus musculus (Mouse).				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
NCB1.TaxId=10090;				
[1]				
SEQUENCE FROM N.A.				
RX MEDLINE=20054576; PubMed=10586086;				
RA Stover C.M., Thiel S., Lynch N.J., Schwaebel W.J.;				
RT "The rat and mouse homologues of MASP-2 and MAP19, components of the				
manan-binding lectin activation pathway of complement.";				
RL J. Immunol. 163:6848-6859(1999).				
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.				
DR EMBL; AJ250369; CAB63701.1; ..				
DR HSSP; P00736; IAPQ.				
DR MGD; MG1:1330832; Masp2.				
DR InterPro; IPRO00152; Asx_hydroxyl.				
DR InterPro; IPRO000859; CUB_domain.				
DR InterPro; IPRO000561; EGF-like.				
DR InterPro; IPRO01881; EGF_Ca.				
DR Pfam; PF00431; CUB; 1.				
DR SMART; SM0042; CUB; 1.				
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.				
DR PROSITE; PS01180; CUB; 1.				
DR PROSITE; PS01186; EGF_2; 1.				
DR PROSITE; PS01187; EGF_CA; 1.				
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat; Signal.				
FT SIGNAL	1	15		POTENTIAL.
FT CHAIN	16	185		MOUSE MAP19
SO SEQUENCE	185 AA;	20973 MW;	72E0190D030A18B9 CRC64;	
	Query Match		76.9%;	Score 183; DB 11; Length 185;
	Best Local Similarity		78.0%;	Fid. No. 4.7e-16;
	Matches	32; Conservative	3; Mismatches	6; Indels
OY				Gaps
Dd	1 TPLGKWPPEVFGRLASPGFPEGYANDDERKWTLPAPGGR 41			
	16 TPLGKWPPEVFGRLASPGFPEKYADHODRWTLPAPGGR 56			
RESULT 5				
090XD5	PRELIMINARY:		PRT: .	190 AA.
090XD5				
01-MAY-2000 (TRENBLREL. 13, Created)				
01-MAY-2000 (TRENBLREL. 13, Last sequence update)				
01-JUN-2002 (TRENBLREL. 21, Last annotation update)				
Mannose binding lectin-associated serine protease-2 precursor.				

OS MAP19.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA;
RX MEDLINE=20054576; PubMed-10586086;
RA Stover C.M., Thiel S., Lynch N.J., Schwaeble W.J.;
RT "The rat and mouse homologues of MASP-2 and Map19, components of the
manan-binding lectin activation pathway of complement.",
RL J. Immunol. 163:6848-6859(1999).
CC -I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; Y19160; CAB65247.1; -.
DR HSSP; P00736; IAPQ.
DR MEROPS; S01.229; -.
DR InterPro; IPRO00152; Asx_hydroxyl.
DR InterPro; IPRO00859; CUB_domain.
DR InterPro; IPRO00561; EGF-like.
DR InterPro; IPRO01881; EGF_Ca.
DR Pfam; PF00431; CUB; 1.
DR SMART; SMO0179; EGF_CA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
KW Repeat; Signal.
FT SIGNAL 1
SQ SEQUENCE 190 AA; 21457 MW; 22BCBAA3F906B010 CRC64;

Query Match 76.9%; Score 183; DB 11; Length 190;
Best Local Similarity 78.0%; Pred. NO. 4,9e-16;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEPVFGRLASPGFPEGVANDGERAWTLTAPPGR 41
II IIIIIIIIIIIIIIIIII:::I:I IIIIIIIII
Db 21 TLGSKWEPEVFGRLVSPGFEPKAYADHQDSMTLTAPPGR 61

RESULT 6
Q9QXD2 PRELIMINARY; PRT; 367 AA.
AC Q9QXD2.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannose binding lectin-associated serine protease-2 precursor
DE (Fragment).
GN MASP2 OR MASP-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6CBA/J;
RX MEDLINE=20054576; PubMed-10586086;
RA Stover C.M., Thiel S., Lynch N.J., Schwaeble W.J.;
RT "The rat and mouse homologues of MASP-2 and Map19, components of the
manan-binding lectin activation pathway of complement.",
RL J. Immunol. 163:6848-6859(1999).
CC -I- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; Y19163; CAB65250.1; -.
DR HSSP; P00736; IAPQ.
DR MEROPS; S01.229; -.
DR MCD; MGI:1330832; Masp2.
DR InterPro; IPRO00152; Asx_hydroxyl.
DR InterPro; IPRO00859; CUB_domain.

DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR004825; Ins/IGF/relax.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB_2.
DR Pfam: PF00084; sush1; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Lectin; Protease;
KW Repeat; signal;
FT SIGNAL 1 20 POTENTIAL.
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 40950 MW; F66FCB29CEAB2B7D CRC64;
Query Match 76.9%; Score 183; DB 11; Length 367;
Local Similarity 78.0%; Pred. No. 1.1e-15;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGPGGEYANDQERRMTLTAPPGYR 41
DB 21 TLGSKWPEVFGRLVSPGPEKYADHDQRSWTLTAPPGYR 61

RESULT 7
ID 092338 PRELIMINARY: PRT; 685 AA.
AC 092338;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Manose-binding lectin associated serine protease-2.
GN MASP2 OR MASP-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=LIVER;
RX MEDLINE=99008558; PubMed=9794427;
RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
Nonaka M., Fujita T.;
RA Two lineages of manose-binding lectin-associated serine protease
(MASP) in vertebrates.";
RA Immunol. 161:4924-4930(1998).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AB009459; BAA34674.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.229; -
DR MGD: MGI:1330832; Masp2.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.

DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Lectin;
KW Protease; Repeat; Serine protease.
SQ SEQUENCE 685 AA; 75490 MW; 4571E1AE0703A70F CRC64;

Query Match 76.9%; Score 183; DB 11; Length 685;
Best Local Similarity 78.0%; Pred. No. 2.2e-15;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGPGGEYANDQERRMTLTAPPGYR 41
DB 16 TLGSKWPEVFGRLVSPGPEKYADHDQRSWTLTAPPGYR 56

RESULT 8
ID 091WPO PRELIMINARY: PRT; 685 AA.
AC 091WPO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to mannan-binding lectin serine protease 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC013893; AAH13893.1; -
DR MEROPS: S01.229; -
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sush1; 2.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00181; EGF; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR PROSITE: PS01187; EGF_CA; UNKNOWN_1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
DR Hydrolyase; Lectin; Protease; Serine protease.
SQ SEQUENCE 685 AA; 75516 MW; F56A6D522BC7099D CRC64;
Query Match 76.9%; Score 183; DB 11; Length 685;
Best Local Similarity 78.0%; Pred. No. 2.2e-15;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 TPLGPKWPEVFGRLASPGPGGEYANDQERRMTLTAPPGYR 41
DB 16 TLGSKWPEVFGRLVSPGPEKYADHDQRSWTLTAPPGYR 56

RESULT 9
ID 09JUS8 PRELIMINARY: PRT; 678 AA.
AC 09JUS8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)


```
DR InterPro: IPR000436; Sushtl_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00084; Sushtl; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Protease;
Repeat; Serine protease.
SEQUENCE 688 AA; 77513 MW; 180DD559EF8B6573 CRC64;
Query Match
Best Local Similarity 42.0%; Score 100; DB 13; Length 688;
Matches 16; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 11 VFGRLASPGFPGPEYANDQERRWTLTAPPGYR 41
DB 29 LFGRISSPGFPRKPSYSDLTMMNNIKVPEGYR 59
RESULT 15
009020 PRELIMINARY; PRT; 212 AA.
ID: 009020;
AC: 009020;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ra-reactive factor serine protease p100 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuitel T., Fellmer P., Neubauer K., Kawakami M., Grundmann A.,
RA Ramadori G.;
RT "The complement activating protease p100 is expressed by hepatocytes
and is induced by IL-6 in vitro and during the acute phase reaction in
vivo.";
ab. Invest. 0:0-0(1997).
I-SIMILARITY: CONTAINS 1 CUB DOMAIN.
EMBL: AF004661; AAB65832.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR Pfam: PF00431; CUB; 2.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00179; EGF_CA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat.
FT NON_TER 1
FT NON_TER 1
FT CHAIN 1
SQ SEQUENCE 212 AA; 24377 MW; 1BB2F52BF0D30A7E CRC64;
Query Match
Best Local Similarity 39.1%; Score 93; DB 11; Length 212;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 11 VFGRLASPGFPGPEYANDQERRWTLTAPPGYR 41
DB 21 MFGQIQSPGYPSPDSSEVTWITVPEGYR 51
RESULT 16
09JUS9 PRELIMINARY; PRT; 701 AA.
ID 09JUS9
AC 09JUS9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannose-binding protein associated serine protease-1 precursor
DE (Fragment).
GN MASP-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=20469449; PubMed=10913141;
RA Wallis R., Dodd R.B.;
RT "Interaction of mannose-binding protein with associated serine
RT proteases: Effects of naturally occurring mutations.";
RL J. Biol. Chem. 275:30962-30969(2000).
CC -I-SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I-SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AJ277423; CAB89695.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.198; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushtl_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; Sushtl; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Protease;
Repeat; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 21
FT CHAIN 22 701
SQ SEQUENCE 701 AA; 79663 MW; B5CFD619D63CE3DD CRC64;
Query Match
Best Local Similarity 39.1%; Score 93; DB 11; Length 701;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
QY 11 VFGRLASPGFPGPEYANDQERRWTLTAPPGYR 41
DB 29 MFGQIQSPGYPSPDSSEVTWITVPEGYR 59
RESULT 17
```

Q96RS4 ID Q96RS4 PRELIMINARY: PRT: 728 AA.
AC Q96RS4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Complement factor MASP-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21378425; PubMed=11485744;
RA Dahl M.R., Thiel S., Matsushita M., Fujita T., Willis A.C.,
RA Christensen T., Vorup-Jensen T., Jensenius J.C.;
RT "Maspl-3 and its association with distinct complexes of the mannan-
binding lectin complement activation pathway.";
RL Immunity 15:127-135(2001).
EMBL: AF284421; AAK84071.1; -
InterPro: IPR000152; Asx_hydroxyl.
InterPro: IPR000859; CUB_domain.
InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sushi; 2.
DR Pfam: PF00089; trypsin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS001010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01187; EGF_CA; UNKNOWN_1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR Hydrolase; Serine protease.
KW SEQUENCE 728 AA; 81860 MW; 09B5297A6C14283A CRC64;
SQ
Query Match 39.1%; Score 93; DB 4; Length 728;
Best Local Similarity 45.2%; Pred. No. 0.0012;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
QY 11 VGRLASPGFPGFYANDQERRWTLTAPGGR 41
DB 27 MEGQIQSPGYPSPSDSEVTWNITVPEGFR 57
T 18
92050 PRELIMINARY: PRT: 733 AA.
92050:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MBL-associated serine protease-3.
GN MASP1 OR MASP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=LIVER;
RA Takahashi M., Miura S., Ishii N., Sugamura K., Shuichi M., Shiro S.,
RA Endo Y., Matsushita M., Fujita T.;
RT "An essential role of MBL-associated serine protease-1 and -3 in
activation of complement by lectin pathway";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB049755; BAB69688.1; -
DR MEROPS: S01.132; -
DR MGD: MGI:88492; Masp1.
DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sushi; 2.
DR Pfam: PF00089; trypsin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS001010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01187; EGF_CA; UNKNOWN_1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 733 AA; 82373 MW; 6B917743118941C1 CRC64;
SQ
Query Match 39.1%; Score 93; DB 11; Length 733;
Best Local Similarity 45.2%; Pred. No. 0.0012;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
QY 11 VGRLASPGFPGFYANDQERRWTLTAPGGR 41
DB 32 MEGQIQSPGYPSPSDSEVTWNITVPEGFR 62
RESULT 19
Q8TD25 ID Q8TD25 PRELIMINARY: PRT: 996 AA.
AC Q8TD25:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE SE26.
GN SE26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY038048; AAK71497.1; -
SQ SEQUENCE 996 AA; 107585 MW; E8012D30969925E CRC64; .
SQ
Query Match 38.4%; Score 91.5; DB 4; Length 996;
Best Local Similarity 40.4%; Pred. No. 0.0028;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;
QY 5 PKW-PEPY-----FGRLASPGFPGFYANDQERRWTLTAPGGR 41
DB 405 PWDSEKPCVCEPCGYRNATGRTIVSPFGNYNNLTCHWLLDAPGGR 456
RESULT 20
Q9UC48 ID Q9UC48 PRELIMINARY: PRT: 16 AA.
AC Q9UC48:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 18.5 kDa tumor marker protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE
RX MEDLINE=96212543; PubMed=8618346;
RA Rasmussen H.H., Orncoft T.F., Wolf H., Celis J.E.;
RT "Towards a comprehensive database of proteins from the urine of

RT patients with bladder cancer."
 RL J. Urol. 155:2113-2119(1996).
 SQ SEQUENCE 16 AA; 1751 MM; 289540537A1B4400 CRC64;

Query Match 36.6%; Score 87; DB 4; Length 16;
 Best Local Similarity 100.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LASPGFPGGYANDQER 30
 ||||||||||||
 DB 1 LASPGFPGGYANDQER 16

Search completed: January 11, 2003, 10:51:43
 Job time : 125 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 10:08:28 ; Search time 65 seconds

(without alignments)
26.162 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238

Sequence: 1 TPLGPKWPEVFGRLASPGF.....GEVANDQERRWLTAPPGRY 41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ed: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	686	1	MAS2_HUMAN
2	93	39.1	699	1	CRAR_HUMAN
3	93	39.1	704	1	CRAR_MOUSE
4	85	35.7	592	1	TOH2_CAEEL
5	81	34.0	705	1	CLR_HUMAN
6	78	32.8	449	1	PCOI_HUMAN
7	74.5	31.3	688	1	CIS_HUMAN
8	74	31.1	468	1	PCOI_MOUSE
9	74	31.1	514	1	PCOI_RAT
10	74	31.1	514	1	UVS2_XENLA
11	74	31.1	1022	1	TLD_BRARE
12	72.5	30.5	695	1	CASP_MESAU
13	69	29.0	986	1	BMP1_HUMAN
14	69	29.0	991	1	BMP1_MOUSE
15	67	28.2	707	1	BMP1_XENLA
16	66	27.7	639	1	BMP1_STRPU
17	66	27.7	914	1	NRP1_CHICK
18	65.5	27.5	503	1	CA19_CHICK
19	65	27.3	1057	1	TLD_DROME
20	64	26.9	925	1	NRP1_RAT
21	63	26.5	922	1	NRP1_RAT
22	63	26.5	923	1	NRP1_HUMAN
23	63	26.5	923	1	NRP1_MOUSE
24	63	26.5	931	1	NRP2_HUMAN
25	63	26.5	1669	1	CA14_HUMAN
26	61	25.6	931	1	NRP2_MOUSE
27	60.5	25.4	1262	1	CA13_CHICK
28	60	25.2	1466	1	CA13_HUMAN
29	59.5	25.0	1024	1	SZ6L_HUMAN
30	59	24.8	636	1	CA13_RAT
31	58	24.4	1464	1	CA13_MOUSE
32	58	24.4	1670	1	CA13_HUMAN
33	57.5	24.2	275	1	TSG6_MOUSE

Hit No. 1 has bidio-
gave play comp. number
to hit no. 3 and 507 results.

ALIGNMENTS

RESULT 1	MAS2_HUMAN	STANDARD:	PRT:	686 AA.	
ID	MAS2_HUMAN	000187: 075754; Q9Y270; Q9B2H0;			
AC	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DI	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)				
DE	(Mannose-binding lectin protein associated serine protease 2) (MASP-2)				
GN	MASP2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=97242412; PubMed=9087411;				
RA	Thiel S., Jensen T.V., Stover C.M., Schwaible W.J., Laursen S.B.,				
RA	Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,				
RA	Reid K.B.M., Jensenius J.C.;				
RT	"A second serine protease associated with mannan-binding lectin that				
RT	activates complement";				
RL	Nature 386:506-510(1997).				
[2]					
RP	SEQUENCE FROM N.A.				
RA	Thiel S., Vorup-Jensen T., Stover C.M., Schwaible W., Laursen S.B.,				
RA	Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,				
RA	Reid K.B.M., Jensenius J.C.;				
RT	"Identification and characterization of a novel protein of the human				
RT	complement system, mannan-binding lectin-associated serine protease-2				
RT	(MASP-2).";				
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.				
[3]					
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99192764; PubMed=10092804;				
RA	Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,				
RA	Jensenius J.C., Schwaible W.J.;				
RT	"Two constituents of the initiation complex of the mannan-binding				
RT	lectin activation pathway of complement are encoded by a single				

34	57.5	24.2	277	1	TSG6_HUMAN	P98066	homo sapien
35	57.5	24.2	1685	1	CAS4_HUMAN	P29400	homo sapien
36	57.5	24.2	2319	1	NTC3_RAT	O91722	rattus norv
37	57	23.9	822	1	PBPA_PSEAE	O07806	pseudomonas
38	56.5	23.7	276	1	TSG6_RABIT	P98065	oryctolagus
39	56.5	23.7	1049	1	CA13_BOVIN	P04258	bos taurus
40	56.5	23.7	1376	1	VCAP_HSEVB	P28920	equine herp
41	56	23.5	597	1	BP10_PARLI	P42674	paracentrot
42	56	23.5	616	1	SPAN_STRPU	P98068	strongyloce
43	56	23.5	1669	1	CA14_MOUSE	P02453	mus musculu.
44	55.5	23.3	1690	1	CA44_HUMAN	P53440	homo sapien
45	55	23.1	241	1	LEQ3_RABIT	P47845	oryctolagus
46	55	23.1	283	1	Q033_CAEEL	O09233	caenorhabdi
47	54.5	22.9	323	1	CC39_CAEEL	O09455	caenorhabdi
48	54.5	22.9	1496	1	CA25_HUMAN	P05997	homo sapien
49	54.5	22.9	2321	1	NTC3_HUMAN	O8um47	homo sapien
50	54	22.7	411	1	PEPT_YERPE	O82i70	yeast
51	54	22.7	443	1	COBB_THEAC	O91x66	thermoplas
52	54	22.7	777	1	PBPA_VIRCH	O9kuc0	v. penicillii
53	54	22.7	928	1	NRP1_XENLA	P28824	xenopus lae
54	54	22.7	1174	1	KPCI_COCHF	O42632	cochliodonu
55	53.5	22.5	476	1	MM10_MOUSE	O55123	mus musculu
56	53.5	22.5	680	1	CA1A_MOUSE	O05306	mus musculu
57	53.5	22.5	891	1	SYV_PYRHO	O58052	pyrococcus
58	53.5	22.5	1328	1	YMT5_YEAST	O04214	saccharomyc
59	53.5	22.5	1328	1	YMT5_YEAST	O04670	saccharomyc
60	53.5	22.5	1755	1	YJ29_YEAST	P47100	saccharomyc

RT structural gene.";
 RL J. Immunol. 162:3481-3490(1999).
 [4]
 RP SEQUENCE FROM N.A.
 RA Park D., Kim B., Baek K., Yoon J.;
 "Structure of human MASP-2 gene.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRYPSIN PROTEASE THAT PRESUMABLY PLAYS AN IMPORTANT ROLE
 IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER
 ACTIVATION IT CLEAVES C4 GENERATING C4a AND C4b.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.sib.ch).

 DR EMBL: Y09926; CAA71059.1; -
 DR EMBL: X98400; CAA67050.1; -
 DR EMBL: Y18287; CAB50735.1; -
 DR EMBL: Y18286; CAB50733.1; -
 DR EMBL: AF321562; AAG50274.1; -
 DR EMBL: AF321558; AAG50274.1; JOINED.
 DR EMBL: AF321559; AAG50274.1; JOINED.
 DR EMBL: AF321560; AAG50274.1; JOINED.
 DR EMBL: AF321561; AAG50274.1; JOINED.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.229; -
 DR Genew: HGNC:6902; MASP2.
 DR MIM: 605102; -
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00084; sush1; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS02040; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KM Hydroxylase; Complement pathway; Serine protease; Protease;
 Glycoprotein; Sush1; Repeat; Signal; EGF-like domain; Hydroxylation.
 KW SIGNAL
 FT CHAIN 1 15
 FT CHAIN 16 686
 FT CHAIN 16 444
 FT CHAIN 445 686
 FT CHAIN 16 137
 FT DOMAIN 138 181
 FT DOMAIN 184 296
 FT DOMAIN 299 362
 FT DOMAIN 365 431
 FT DOMAIN 445 686
 FT ACT_SITE 483 483
 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 532 532 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 633 633 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT MOD_RES 158 158 HYDROXYLATION (POTENTIAL).
 FT DISULFID 142 156 POTENTIAL.
 FT DISULFID 142 156 POTENTIAL.
 FT DISULFID 152 165 POTENTIAL.
 FT DISULFID 167 180 POTENTIAL.
 FT DISULFID 184 211 POTENTIAL.
 FT DISULFID 241 259 POTENTIAL.
 FT DISULFID 300 348 POTENTIAL.
 FT DISULFID 328 361 POTENTIAL.
 FT DISULFID 366 412 POTENTIAL.
 FT DISULFID 396 430 POTENTIAL.
 FT DISULFID 434 552 INTERCHAIN (POTENTIAL).
 FT DISULFID 598 618 POTENTIAL.
 FT DISULFID 629 660 POTENTIAL.
 FT CONFLICT 361 362 MISSING (IN REF. 3).
 FT CONFLICT 371 371 D -> Y (IN REF. 4).
 FT CONFLICT 372 372 L -> LCS (IN REF. 3).
 FT CONFLICT 442 442 G -> E (IN REF. 4).
 FT CONFLICT 447 447 G -> E (IN REF. 4).
 FT CONFLICT 461 462 MISSING (IN REF. 3).
 FT CONFLICT 473 473 L -> LIL (IN REF. 3).
 SQ SEQUENCE 686 AA; 75685 MW; 4E34DED159448A2A CRC64;
 Query Match 100.0%; Score 238; DB 1; Length 686;
 Best Local Similarity 100.0%; Pred No. 1.8e-20;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TPGLPKWPEVFGRLASPGPGGEYANDQERRWLTAPGYR 41
 16 TPGLPKWPEVFGRLASPGPGGEYANDQERRWLTAPGYR 56
 Db 16 TPGLPKWPEVFGRLASPGPGGEYANDQERRWLTAPGYR 56
 RESULT 2
 ID CRAR_HUMAN STANDARD; PRT; 699 AA.
 AC P48740; O95570; O9UFE09;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement-activating component of Ra-reactive factor precursor
 DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RaRf)
 DE (Mannan-binding lectin serine protease 1) (Mannose-binding protein
 DE associated serine protease) (MASP-1).
 GN MASP1 OR CRARF OR CRARF1 OR PRSS5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Liver;
 RC MEDLINE=94059062; PubMed=8240317;
 RA Takada F., Takayama Y., Hatsuse H., Kawakami M.;
 RT "A new member of the Cls family of complement proteins found in a
 RT bactericidal factor, Ra-reactive factor, in human serum.";
 RL Biochem. Biophys. Res. Commun. 196:1003-1009(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Fetal liver;
 RC MEDLINE=94289349; PubMed=8018603;
 RA Sato T., Endo Y., Matsushita M., Fujita T.;
 RT "Molecular characterization of a novel serine protease involved in
 RT activation of the complement system by mannose-binding protein.";
 RL Int. Immunol. 6:665-669(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=97079701; PubMed=8921412;
 RA Endo Y., Sato T., Matsushita M., Fujita T.;
 RT "Exon structure of the gene encoding the human mannose-binding
 RT protein-associated serine protease light chain: comparison with

RT complement C1r and C1s genes.";
RN Int. Immunol. 8:1355-1358(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-99402590; PubMed-10475605;
RA Takayama Y., Takada F., Nowatari M., Kawakami M., Matsu-ura N.;
RT "Gene structure of the p100 serine-protease component of the human Ra-
reactive factor.";
RL Mol. Immunol. 36:505-514(1999).
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDIAL RA-REACTIVE FACTOR RARF
WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
(CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)
LINKED BY A DISULFIDE BOND.
CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17525; BAA04477.1; -
DR EMBL; D28593; BAA05928.1; -
DR EMBL; D61695; BAA34864.1; -
DR EMBL; AB010822; BAA34864.1; JOINED.
DR EMBL; AB010813; BAA34864.1; JOINED.
DR EMBL; AB010814; BAA34864.1; JOINED.
DR EMBL; AB010815; BAA34864.1; JOINED.
DR EMBL; AB010816; BAA34864.1; JOINED.
DR EMBL; AB010817; BAA34864.1; JOINED.
DR EMBL; AB010818; BAA34864.1; JOINED.
DR EMBL; AB010819; BAA34864.1; JOINED.
DR EMBL; AB010820; BAA34864.1; JOINED.
DR EMBL; AB010821; BAA34864.1; JOINED.
DR EMBL; D61690; BAA34864.1; JOINED.
DR EMBL; D61691; BAA34864.1; JOINED.
DR EMBL; D61692; BAA34864.1; JOINED.
DR EMBL; D61693; BAA34864.1; JOINED.
DR EMBL; D61694; BAA34864.1; JOINED.
DR EMBL; AB007617; BAA89206.1; -
DR EMBL; AB007602; BAA89206.1; JOINED.
DR EMBL; AB007603; BAA89206.1; JOINED.
DR EMBL; AB007604; BAA89206.1; JOINED.
DR EMBL; AB007605; BAA89206.1; JOINED.
DR EMBL; AB007606; BAA89206.1; JOINED.
DR EMBL; AB007607; BAA89206.1; JOINED.
DR EMBL; AB007608; BAA89206.1; JOINED.
DR EMBL; AB007609; BAA89206.1; JOINED.
DR EMBL; AB007610; BAA89206.1; JOINED.
DR EMBL; AB007611; BAA89206.1; JOINED.
DR EMBL; AB007612; BAA89206.1; JOINED.
DR EMBL; AB007613; BAA89206.1; JOINED.
DR EMBL; AB007614; BAA89206.1; JOINED.
DR EMBL; AB007615; BAA89206.1; JOINED.
DR EMBL; AB007616; BAA89206.1; JOINED.
DR HSSP; P00736; IAPQ.
DR MEROPS; S01.198; -
DR Genew; HGNC:6901; MASPL.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001254; Ser. protease_Try.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 2.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00431; CUB; 2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Complement pathway; Serine protease; Protease;
KW Glycoprotein; Sush1; Repeat; Signal; EGF-like domain; Hydroxylation.
FT SIGNAL 1 19
FT CHAIN 20 699
FT FT
FT CHAIN 20 448
FT CHAIN 449 699
FT DOMAIN 20 138
FT DOMAIN 139 182
FT DOMAIN 185 297
FT DOMAIN 300 363
FT DOMAIN 366 433
FT DOMAIN 449 699
FT ACT_SITE 490 490
FT ACT_SITE 532 532
FT ACT_SITE 646 646
FT MOD_RES 159 159
FT DISULFID 73 91
FT DISULFID 143 157
FT DISULFID 153 166
FT DISULFID 168 181
FT DISULFID 185 212
FT DISULFID 242 260
FT DISULFID 301 349
FT DISULFID 329 362
FT DISULFID 367 414
FT DISULFID 397 432
FT DISULFID 436 572
FT DISULFID 614 631
FT DISULFID 642 672
FT CARBOHYD 49 49
FT CARBOHYD 178 178
FT CARBOHYD 385 385
FT CARBOHYD 407 407
FT CONFLICT 235 235
FT CONFLICT 285 285
FT CONFLICT 499 499
FT CONFLICT 499 499
FT CONFLICT 527 527
FT CONFLICT 543 543
FT CONFLICT 552 552
FT CONFLICT 643 643
SQ SEQUENCE 699 AA; 79258 MW; ADD9697AE6AB01B5 CRC64;

Query Match 39.1%; Score 93; DB 1; Length 699;
Best Local Similarity 45.2%; Pred. No. 0.0016;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

OY 11 VEGRLASPGFGEYANQDERWTLTAPPGYR 41
DB 27 MFGQISPGYPSDYSPDSEVTWNTVDPGFR 57

ID	CRAR_MOUSE	STANDARD:	PRT:	704 AA.
AC	P96064.			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Complement-activating component of Ra-reactive factor precursor (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RARF)			
DE	(Mannan-binding lectin serine protease 1).			
OS	Mus musculus (Mouse).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RN	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-BALB/c; TISSUE=Liver;			
RC	MEDLINE=94179811; PubMed=8133044;			
RX	Takayama Y., Takada F., Takahashi A., Kawakami M.,			
RX	A 100-kDa protein in the C4-activating component of Ra-reactive			
RT	factor is a new serine protease having module organization similar to			
RT	Cir and C1s.			
RL	J. Immunol. 152:2308-2316(1994).			
RN	[2]			
RN	SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-BALB/c; TISSUE=Liver;			
RC	MEDLINE=93176166; PubMed=8439319;			
RA	Takahashi A., Takayama Y., Hatsuue H., Kawakami M.;			
RT	"Presence of a serine protease in the complement-activating component			
RT	of the complement-dependent bactericidal factor, RARF, in mouse			
RT	serum."			
RL	Biochem. Biophys. Res. Commun. 190:681-687(1993).			
CC	-1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF			
CC	WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY			
CC	CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT			
CC	CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE			
CC	C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.			
CC	-1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT			
CC	(CARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.			
CC	CARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (29)			
CC	LINKED BY A DISULFIDE BOND.			
CC	-1- TISSUE SPECIFICITY: LIVER.			
CC	-1- DOMAIN: CRAR HAS A MODULE ORGANIZATION SIMILAR TO CIR AND C1S.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC	-1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D16492; BAA03944.1; ..			
DR	HSSP; P00736; IAP0.			
DR	MEPROS; S01.198; ..			
DR	MCD; MGI:88492; Maspl.			
DR	InterPro; IPR000152; Asx_hydroxyl.			
DR	InterPro; IPR000859; CUB_domain.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001254; Ser_protease_Try.			
DR	InterPro; IPR000436; Sushi_SCR_CCP.			
DR	Pfam; PF00084; sushi; 2.			
DR	Pfam; PF00089; trypsin; 1.			
DR	Pfam; PF00431; CUB; 2.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00032; CCP; 2.			

Query Match	Best Local Similarity	Score	DB 1:	Length	704:
Matches 14; Conservative	45.2%	7; Mismatches 10; Indels 0; Gaps 0;			
Qy 11 VEGRLASPGCEGVANDQERRWTLTAAPQYR 41	11	11	11	11	11
Db 32 MEGQIQSPGYDPDPSYDSEVTWNTIVPEGR 62	32	32	32	32	32
RESULT 4					
TOH2_CAEEL	STANDARD:	PRT:	592 AA.		
AC P98060:					
DT 01-FEB-1996 (Rel. 33, Created)					
DT 16-OCT-2001 (Rel. 40, Last sequence update)					
DT 16-OCT-2001 (Rel. 40, Last annotation update)					
DE Zinc metalloproteinase toh-2 precursor (EC 3.4.24.-)					
GN TOH-2 OR R151.5.					
OS Caenorhabditis elegans.					
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoides;					
OC Rhabdilitidae; Pelodierinae; Caenorhabditis.					
NCBI TaxID=6239;					
RM [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=Bristol NZ;					
RL Fulton L.,					
Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.					

CC -1- PTM: C-TERMINALLY PROCESSED AT MULTIPLE POSITIONS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L33799; AAA61949.1; ALT_SEQ.
 CC EMBL: AB008549; BAA23281.1; -
 CC EMBL: AF053356; AAC78800.1; -
 CC EMBL: AF083655; AAD16041.1; -
 CC Genes: HGNC:8738; PCOLCE.
 CC MIM: 600270; -
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR001134; Netrin_C.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF01759; NTR; 1.
 CC SMART: SM00042; CUB; 2.
 CC PROSITE: PS01180; CUB; 2.
 CC Glycoprotein: Repeat; Signal.
 CC SIGNAL 1 25 POTENTIAL.
 CC CHAIN 26 449 PROCOLLAGEN C-PROTEINASE ENHANCER
 CC PROTEIN.
 CC FT DOMAIN 37 149 CUB 1.
 CC FT 159 273 CUB 2.
 CC FT DOMAIN 329 438 NTR.
 CC FT SITE 287 288 CLEAVAGE.
 CC FT SITE 288 289 CLEAVAGE.
 CC FT SITE 293 294 CLEAVAGE.
 CC FT SITE 299 300 CLEAVAGE.
 CC SITE 303 304 CLEAVAGE.
 CC CARBOHYD 29 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SO SEQUENCE 449 AA; 47972 MW; 3D88430158648796 CRC64;

Query Match 32.8%; Score 78; DB 1; Length 449;
 Best Local Similarity 37.8%; Pred. No. 0.056;
 Matches 17; Conservative 5; Mismatches 13; Indels 10; Gaps 1;

QY 5 PKRPEPVF-----GRLASGFGGEVANYOERRWTLTAPP 39
 28 PNTRPVFLCGDYKSGESGIVASEGPNLYPKKECIWITVPEG 72

CIS_HUMAN
 ID CIS_HUMAN STANDARD: PRT: 688 AA.
 AC P09871; Q9UM14; Q9UCU7; Q9UCU8; Q9UCU9; Q9UCV1; Q9UCV2;
 AC Q9UCV3; Q9UCV4; Q9UCV5;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement C1s component precursor (EC 3.4.21.42) (C1 esterase).
 GN C1S.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89017187; PubMed-2459702;
 RA Kusumoto H., Hirose S., Salier J.P., Hagen F.S., Kurechi K.;
 RT "Human genes for complement components C1r and C1s in a close
 RT tail-to-tail arrangement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7307-7311(1988).
 RN [12]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Liver;

RX MEDLINE-88082788; PubMed-3500856;
 RA McKinnon C.M., Carter P.E., Smyth S.J., Dunbar B., Fothergill J.E.;
 RT "Molecular cloning of cDNA for human complement component C1s. The
 RT complete amino acid sequence.";
 RL Eur. J. Biochem. 169:547-553(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88163522; PubMed-2831944;
 RA Tosti M., Duponchel C., Meo T., Julier C.;
 RT "Complete cDNA sequence of human complement C1s and close physical
 RT linkage of the homologous genes C1s and C1r.";
 RL Biochemistry 26:8516-8524(1987).
 RN [4]
 RP SEQUENCE OF 1-329 FROM N.A.
 RC TISSUE=Peripheral blood leukocytes;
 RX MEDLINE-99008558; PubMed-9794427;
 RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
 RA Nonaka M., Fujita T.;
 RT "Two lineages of mannose-binding lectin-associated serine protease
 RT (MASP) in vertebrates.";
 RL J. Immunol. 161:4924-4930(1998).
 RN [5]
 RP SEQUENCE OF 291-688 FROM N.A.
 RX MEDLINE-90040704; PubMed-2553984;
 RA Tosti M., Duponchel C., Meo T., Couture-Tosi E.;
 RT "Complement genes C1r and C1s feature an intronless serine protease
 RT domain closely related to haptoglobin.";
 RL J. Mol. Biol. 208:709-714(1989).
 RN [6]
 RP SEQUENCE OF 16-61; 168-219; 287-334 AND 384-445.
 RX MEDLINE-86164350; PubMed-3007145;
 RA Spycher S.E., Nick H., Rickli E.E.;
 RT "Human complement component C1s. Partial sequence determination of
 RT the heavy chain and identification of the peptide bond cleaved during
 RT activation.";
 RL Eur. J. Biochem. 156:49-57(1986).
 RN [7]
 RP SEQUENCE OF 438-500; 503-534; 542-601; 617-623 AND 626-656.
 RX MEDLINE-84104122; PubMed-6362661;
 RA Carter P.E., Dunbar B., Fothergill J.E.;
 RT "The serine proteinase chain of human complement component C1s.
 RT Cyanogen bromide cleavage and N-terminal sequences of the
 RT fragments.";
 RL Biochem. J. 215:565-571(1983).
 RN [8]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Plasma;
 RX MEDLINE-91308095; PubMed-1854725;
 RA Ily C., Thieles N.M., Gagnon J., Arlaud G.J.;
 RT "Effect of lactoperoxidase-catalyzed iodination on the C4(2+)-
 RT dependent interactions of human C1s. Location of the iodination
 RT sites.";
 RL Biochemistry 30:7135-7141(1991).
 RN [9]
 RP DISULFIDE BONDS.
 RX MEDLINE-91175725; PubMed-2007122;
 RA Hess D., Schaller J., Rickli E.E.;
 RT "Identification of the disulfide bonds of human complement C1s.";
 RL Biochemistry 30:2827-2833(1991).
 RN [10]
 RP PARTIAL SEQUENCE, AND 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
 RX MEDLINE-95288736; PubMed-7797774;
 RA Rossi V., Gabordiau C., Lacroix M., Ulrich J., Fontecilla-Camps J.C.,
 RA Gagnon J., Arlaud G.J.;
 RT "Structure of the catalytic region of human complement protease C1s:
 RT study by chemical cross-linking and three-dimensional homology
 RT modeling.";
 RL Biochemistry 34:7311-7321(1995).
 RN [11]
 RP DISEASE.
 RX PubMed-11390518;
 RA Dragon-Durey M.-A., Quartier P., Fremiaux-Bacchi V., Blouin J.,
 RA de Barce C., Prieur A.-M., Weiss L., Fridman W.-H.;

RT "Molecular basis of a selective C1s deficiency associated with early
 RT onset multiple autoimmune diseases." ;
 RL J. Immunol. 166:7612-7616(2001).
 CC -1- FUNCTION: C1S B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1O
 CC AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
 CC OF THE COMPLEMENT SYSTEM. C1R ACTIVATES C1S SO THAT IT CAN, IN
 CC TURN, ACTIVATE C2 AND C4.
 CC -1- CATALYTIC ACTIVITY: CLEAVES COMPONENT C4 TO C4A AND C4B, AND
 CC COMPONENT C2 TO C2A AND C2B.
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1O, R
 CC AND S IN THE MOLAR RATION OF 1:2:2. ACTIVATED C1S IS AN DISULFIDE-
 CC LINKED HETERODIMER OF AN HEAVY CHAIN AND A LIGHT CHAIN.
 CC -1- DISEASE: Defects in C1S are the cause of selective C1s deficiency,
 CC that is associated with early onset multiple autoimmune diseases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X06596; CAA29817.1; -;
 CC EMBL: J04080; AAA51852.1; -;
 CC EMBL: M18767; AAA51853.1; -;
 CC EMBL: AB009076; BAA86864.1; -;
 CC PIR: A40496; C1HUS.
 CC PIR: S00224; S00224.
 CC MEROPS: S01.193; -;
 CC SWISS-2PAGE: P09871; HUMAN.
 CC GeneW: HGNC:1247; C1S.
 CC MIM: 120580; -;
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR001314; Chymotrypsin.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001254; Ser_protease_Try.
 CC InterPro: IPR000436; Sushi_SCR_CCP.
 CC Pfam: PF00008; EGF.1.
 CC Pfam: PF00084; sushi.2.
 CC Pfam: PF00089; trypsin.1.
 CC Pfam: PF00431; CUB.2.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00032; CCP.2.
 CC SMART: SM00042; CUB.2.
 CC SMART: SM00179; EGF_CA.1.
 CC SMART: SM00020; TRYP_SPE.1.
 CC SMART: PS00010; ASX_HYDROXYL.1.
 CC PROSITE: PS00180; CUB.2.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC PROSITE: PS01186; EGF_2; FALSE_NEG.
 CC PROSITE: PS01187; EGF_CA.1.
 CC PROSITE: PS00240; TRYPSIN_DOM.1.
 CC PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 CC PROSITE: PS00135; TRYPSIN_SER.1.
 CC Complement pathway; Plasma; Glycoprotein; Serine protease; Hydrolase;
 CC Hydroxylation; Sushi; Repeat; Signal; EGF-like domain;
 CC Calcium-binding.
 CC SIGNAL 1 15
 FT CHAIN 16 437 COMPLEMENT C1S HEAVY CHAIN.
 FT CHAIN 438 688 COMPLEMENT C1S LIGHT CHAIN.
 FT DOMAIN 16 130 CUB 1.
 FT DOMAIN 131 172 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 175 290 CUB 2.
 FT DOMAIN 293 355 SUSHI 1.
 FT DOMAIN 358 422 SUSHI 2.
 FT DOMAIN 438 688 SERINE PROTEASE.

FT ACT_SITE 475 475 CHARGE RELAY SYSTEM.
 FT ACT_SITE 529 529 CHARGE RELAY SYSTEM.
 FT ACT_SITE 632 632 CHARGE RELAY SYSTEM.
 FT MOD_RES 149 149 HYDROXYLATION (PROBABLE).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .).
 FT DISULFID 65 83
 FT DISULFID 135 147
 FT DISULFID 143 156
 FT DISULFID 158 171
 FT DISULFID 175 202
 FT DISULFID 234 251
 FT DISULFID 294 341
 FT DISULFID 321 354
 FT DISULFID 359 403
 FT DISULFID 386 421
 FT DISULFID 425 549
 FT DISULFID 595 618
 FT DISULFID 628 659
 FT DISULFID 294 294 C -> K (IN REF. 6).
 FT CONFLICT 513 513 G -> GG (IN REF. 5).
 FT CONFLICT 573 573 T -> A (IN REF. 7).
 FT CONFLICT 645 646 TK -> GR (IN REF. 7).
 SQ SEQUENCE 688 AA; 76684 MW; 85522647A4C47205 CRC64;

 QY 7 WPEP-VFGRLASPFGEVANDERRMTLTAPPGY 40
 Db 14 YAEPTWGETLSPNYQPAVPSEVKSMDEVPGEY 48

 RESULT 8
 ID PCOL_MOUSE STANDARD; PRT; 468 AA.
 AC 061398; O35113;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I
 DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-
 DE proteinase enhancer protein) (P14).
 GN PCOLCE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=95014462; PubMed=7523404;
 RA Takahara K., Kessler E., Biniashov L., Brusel M., Eddy R.L.,
 RA Jani-Salt S., Shows T.B., Greenspan D.S.;
 RT "Type I procollagen COOH-terminal proteinase enhancer protein:
 RT identification, primary structure, and chromosomal localization of the
 RT cognate human gene (PCOLCE).";
 RL J. Biol. Chem. 269:26280-26285(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Heart;
 RA Hirahara I., Syoutuda K., Harada K., Tomita M., Urakami K., Terai H.,
 RA Morisaki N., Saito Y.;
 RT "Smooth muscle cell derived procollagen C-protease enhancer protein.";
 RL Cell Struct. Funct. 21:662-662(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=CNS;
 RX MEDLINE=91225708; PubMed=2027019;
 RA Iecan E., Zelenika D., Iaine M.C., Rhyner T., Pessac B.;
 RT "Isolation of a novel cDNA corresponding to a transcript expressed in
 RT the choroid plexus and leptomeninges.";
 RL J. Neurochem. 56:2133-2138(1991).

```

RN [4]
RP CHARACTERIZATION.
RX MEDLINE=90092088; PubMed=2689170;
RA Kessler E., Adar R.;
RT "Type I procollagen C-proteinase from mouse fibroblasts. Purification
RL and demonstration of a 55-kDa enhancer glycoprotein."
RL Eur. J. Biochem. 186:115-121(1989).
RN [5]
RP TISSUE SPECIFICITY.
RC STRAIN=ICR;
RX MEDLINE=91076903; PubMed=2256940;
RA Kessler E., Mould A.P., Humes D.J.S.;
RT "Procollagen type I C-proteinase enhancer is a naturally occurring
RL connective tissue glycoprotein."
RL Biochem. Biophys. Res. Commun. 173:81-86(1990).
RN [6]
RP CARBOHYDRATE-LINKAGE SITE ASN-28.
RA Kessler E.;
RT "Unpublished observations (Jan-2000).
RL -1- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN INTERSTITIAL CONNECTIVE TISSUES
CC LIKE TENDONS, CALVARIA, SKIN AND AT A LOWER LEVEL IN HEART AND
CC SKELETAL MUSCLE.
CC -1- PTM: PROCESSED FROM A 55 KDA FORM TO 36 AND 34 KDA FORMS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC IN POSITION 391 AND 445.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB008548; BAA23280.1; -
CC EMBL: X57337; CAA0612.1; ALT_FRAME.
CC MGd; MG1:105099; PcoIce.
CC InterPro: IPR000859; CUB_domain.
CC InterPro: IPR001134; Netrin_C.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF01759; NTR; 1.
CC SMART: SM00042; CUB; 2.
CC PROSITE: PS01180; CUB; 2.
CC -1- glycoprotein; Repeat; Signal.
CC SIGNAL 1 24
CC CHAIN 1 25 468
CC FT DOMAIN 36 148
CC FT DOMAIN 158 272
CC FT DOMAIN 352 461
CC FT CARBOHYD 28 28
CC FT CARBOHYD 454 454
CC FT CONFLICT 178 178
CC FT CONFLICT D -> Y (IN REF. 3) (POTENTIAL).
CC SO SEQUENCE 468 AA; 50168 MW; F9B5558147FAAB CRC64;

Query Match 31.1%; Score 74; DB 1; Length 468;
Best Local Similarity 35.6%; Pred. No. 0.17;
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;
QY 5 PKPPEYF-----GLASPGFPGGEVANDQERRWITLAPPG 39
ID 1 111 1 111 111 1 111 111 1
DB 27 PNTYRPVFLCGDVTGESGVASEGPNLYPKKCIWITIVGEV 71

RESULT 9
PCOL_RAT
ID PCOL_RAT
AC 008628; STANDARD; PRT; 468 AA.

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I
DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-
DE proteinase enhancer protein).
GN PCOLCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mistar;
RX MEDLINE=97447750; PubMed=9303490;
RA Ogata I., Auster A.S., Matsui A., Greenwel P., Ceerts A., D'Amico T.,
RA Fujiwara K., Kessler E., Rojkind M.;
RT "Up-regulation of type I procollagen C-proteinase enhancer protein
RL messenger RNA in rats with CCl4-induced liver fibrosis."
RL Hepatology 26:611-617(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mistar; TISSUE=Aorta;
RA Hirahara I., Syoutuda K., Harada K., Tomita M., Urakami K., Terai H.,
RA Morisaki N., Saito Y.;
RT "Smooth muscle cell derived procollagen C-protease enhancer protein."
RL Cell Struct. Funct. 21:662-662(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344;
RA Masuda M., Igarashi H., Kano M., Yoshikura H.;
RT "Protein integration into the procollagen C-proteinase enhancer
RL protein gene and its effects in cultured rat fibroblasts revealed by
RL an excisable 'hit-and-run' retroviral vector."
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=91076903; PubMed=2256940;
RA Kessler E., Mould A.P., Humes D.J.S.;
RT "Procollagen type I C-proteinase enhancer is a naturally occurring
RL connective tissue glycoprotein."
RL Biochem. Biophys. Res. Commun. 173:81-86(1990).
CC -1- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN COLLAGEN-RICH
CC TISSUES, ESPECIALLY TENDON. ALSO EXPRESSED IN CORNEA AND STERNA.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U94710; AAB93478.1; -
CC EMBL: AB008534; BAA23217.1; -
CC EMBL: AF016503; AAD01592.1; -
CC EMBL: AF016506; AAD01598.1; -
CC InterPro: IPR000859; CUB_domain.
CC InterPro: IPR001134; Netrin_C.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF01759; NTR; 1.
CC SMART: SM00042; CUB; 2.
CC PROSITE: PS01180; CUB; 2.
CC -1- glycoprotein; Repeat; Signal.
CC SIGNAL 1 24
CC CHAIN 1 25 468
CC FT DOMAIN 36 148
CC FT DOMAIN 158 272
CC FT DOMAIN 352 461
CC FT CARBOHYD 28 28
CC FT CARBOHYD 454 454
CC FT CONFLICT 178 178
CC FT CONFLICT D -> Y (IN REF. 3) (POTENTIAL).
CC SO SEQUENCE 468 AA; 50168 MW; F9B5558147FAAB CRC64;

```

FT DOMAIN 158 272 CUB 2.
FT DOMAIN 352 461 NTR.
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 468 AA: 50185 MW: B4AA1C151323969B CRC64;
Query Match 31.1%; Score 74; DB 1; Length 468;
Best Local Similarity 35.6%; Pred. No. 0.17;
Matches 16; Conservative 6; Mismatches 13; Indels 10; Gaps 1;
QY 5 PKMPEPYF-----GLASPGFGEVANDERMTITAPGR 39
DB 27 PNTRPVFLCGDVTEGSGVYASEGPNLYPKKCIWITVPEG 71
RESULT 10
UWS2_XENLA STANDARD; PRT; 514 AA.
AC P42664;
01-NOV-1995 (Rel. 32, Created)
11-NOV-1997 (Rel. 35, Last sequence update)
05-JUN-2002 (Rel. 41, Last annotation update)
Embryonic protein UWS.2 precursor (EC 3.4.24.-).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97228908; PubMed=9074934;
RA Katsuragi C., Maeda R., Yamashita K., Mita K., Sargent T.D.,
RA Yasumasa S.;
RT "Molecular cloning of Xenopus hatching enzyme and its specific
RT expression in hatching gland cells.";
RL Int. J. Dev. Biol. 41:19-25(1997).
RN [2]
RP SEQUENCE OF 196-514 FROM N.A.
RA MEDLINE=90108413; PubMed=1688538;
RA Sato S.M., Sargent T.D.;
RT "Molecular approach to dorsoanterior development in Xenopus laevis.";
RL Dev. Biol. 137:135-141(1990).
RN [3]
RP CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: EXCLUSIVELY IN THE ANTERIOR NEURAL FOLD OF
CC NEURULA STAGE EMBRYOS. BY THE TAILBUD STAGE, THE PROTEIN IS
CC LOCALIZED IN SPECIALIZED CEPHALIC ECTODERM, IN A REGION PROBABLY
CC CORRESPONDING TO THE HATCHING GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D89633; BAA14003.1; -
DR EMBL: M27162; AAA49980.1; -
DR HSSP: P07384; 1IAE.
DR MEROPS: M12.014; -
DR InterPro: IPR001506; Astacin.
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR000130; Zn_MTPeptidase.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF01400; Astacin; 1.
DR PRINTS: PR00480; ASTACIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00235; ZINC; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS01180; CUB; 2.
KW Hydrolase; Metalloprotease; Zinc; Repeat; Signal.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 514 EMBRYONIC PROTEIN UVS.2.
FT DOMAIN 288 400 CUB 1.
FT DOMAIN 402 513 CUB 2.
FT METAL 186 166 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 187 187 BY SIMILARITY.
FT METAL 190 190 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 196 196 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 514 AA: 56844 MW: DELFCB599014D986 CRC64;
Query Match 31.1%; Score 74; DB 1; Length 514;
Best Local Similarity 44.0%; Pred. No. 0.19;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
QY 17 SPFGFGEVANDERMTITAPGR 41
DB 415 SPVPGVYTTNCTWITAPAGFK 439
RESULT 11
TLD_BRARE STANDARD; PRT; 1022 AA.
ID TLD_BRARE
AC 057460;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dorsal-ventral patterning tolloid protein precursor (EC 3.4.24.-)
DE (Mini fin protein).
GN TOLLOID OR TLD OR MPN.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC TISSUE=Gastrula;
RA MEDLINE=98057457; PubMed=9395394;
RA Blader P., Rastegar S., Fischer N., Straehle U.;
RT "Cleavage of the BMP-4 antagonist chordin by zebrafish Tolloid.";
RL Science 278:1937-1940(1997).
RN [2]
RP FUNCTION, AND TISSUE SPECIFICITY.
RA MEDLINE=99307076; PubMed=10375503;
RA Connors S.A., Trout J., Ekker M., Mullins M.C.;
RT "The role of tolloid/minf in dorsoventral pattern formation of the
RT zebrafish embryo.";
RL Development 126:3119-3130(1999).
CC -1- FUNCTION: Required for patterning ventral tissues of the tail. May
CC increase bone morphogenetic protein (BMP) activity at the end of
CC gastrulation by proteolytic cleavage of chordin and release of BMP
CC from inactive complexes.
CC -1- TISSUE SPECIFICITY: During gastrulation, accumulates around the
CC closing blastopore with greater expression ventrally. At the
CC animal pole, expressed in the ectoderm flanking the anterior
CC neural plate. At the 10-somite stage, expressed in the developing
CC tailbud and cranial neural crest. At the 20-somite stage, also
CC expressed in the hematopoietic system.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 CUB DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@lsb-sib.ch.

```

CC -----
CC EMBL: AF027596; AAC60304.1; -
CC DR HSSP: P35555; 1EMN.
CC DR MEROPS: M12.016; -
CC ZFIN: ZDB-GENE-990415-265; tollold.
CC InterPro: IPR001506; Asx_hydroxyl.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000859; CUB_domain.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR000130; Zn_Mrppeptidase.
CC Pfam: PF00008; EGF; 2.
CC Pfam: PF00431; CUB; 5.
CC Pfam: PF01400; Astacin; 1.
CC PRINTS: PR00480; ASTACIN.
CC SMART: SM00042; CUB; 5.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00179; EGF_CA; 2.
CC SMART: SM00235; Zmnc; 1.
CC PROSITE: PS00010; ASX_HYDROXYL; 2.
CC PROSITE: PS01180; CUB; 5.
CC PROSITE: PS00022; EGF_1; FALSE_NEG.
CC PROSITE: PS01186; EGF_2; 2.
CC PROSITE: PS01187; EGF_CA; 2.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Developmental protein: Protease; Metalloprotease; Zinc;
KM Metal-binding; Calcium; EGF-like domain; Repeat; Signal; Glycoprotein;
KM Zymogen.
FT SIGNAL 1 32 POTENTIAL.
FT PROPEP 33 156 POTENTIAL.
FT CHAIN 157 1022 DORSAL-VENTRAL PATTERNING TOLLOID
FT DOMAIN 157 357 METALLOPROTEASE (BY SIMILARITY).
FT DOMAIN 358 470 CUB 1.
FT DOMAIN 471 583 CUB 2.
FT DOMAIN 583 624 EGF-LIKE 1, CALCIUM-BINDING
(POTENTIAL).
FT DOMAIN 627 739 CUB 3.
FT DOMAIN 739 779 EGF-LIKE 2, CALCIUM-BINDING
(POTENTIAL).
FT DOMAIN 783 895 CUB 4.
FT DOMAIN 896 1012 CUB 5.
FT METAL 249 249 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 250 250 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 253 253 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 259 259 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 358 384 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 411 433 BY SIMILARITY.
FT METAL 433 497 BY SIMILARITY.
FT METAL 497 546 BY SIMILARITY.
FT METAL 524 546 BY SIMILARITY.
FT METAL 546 599 BY SIMILARITY.
FT METAL 599 608 BY SIMILARITY.
FT METAL 608 623 BY SIMILARITY.
FT METAL 623 653 BY SIMILARITY.
FT METAL 653 702 BY SIMILARITY.
FT METAL 702 743 BY SIMILARITY.
FT METAL 743 754 BY SIMILARITY.
FT METAL 754 763 BY SIMILARITY.
FT METAL 763 778 BY SIMILARITY.
FT METAL 778 809 BY SIMILARITY.
FT METAL 809 858 BY SIMILARITY.
FT METAL 858 926 BY SIMILARITY.
FT METAL 926 975 BY SIMILARITY.
FT METAL 975 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 129 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 178 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 368 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 399 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1022 AA; 115536 MW; A68CA1D0E41793F9 CRC64;

```

Query Match 31.1%; Score 74; DB 1; Length 1022;
 Best Local Similarity 41.4%; Pred. No. 0.38;
 Matches 12; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

OY 13 GRASPPEPEYANDQERRMTLPAPGYR 41
 DB 792 GTISSPMPDKIPSRKECTWIDITATPGHR 820

RESULT 12

CASP_MESAU STANDARD; PRT; 695 AA.
 AC P15156.
 DT 01-APR-1990 (rel. 14, Created)
 DT 01-APR-1990 (rel. 14, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Calcium-dependent serine proteinase precursor (EC 3.4.21.-) (CASP).
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxId=10036;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-50 AND 446-472.
 RC TISSUE=Fibroblast;
 RX MEDLINE=89325606; PubMed=2753140;
 RA Kinoshita H., Sakiyama H., Tokunaga K., Imajob-Ohmi S., Hamada Y.,
 RA Isono K., Sakiyama S.;
 RT "Complete primary structure of calcium-dependent serine proteinase
 capable of degrading extracellular matrix proteins.";
 RL FEBS Lett. 250:411-415(1989).
 CC - FUNCTION: CAPABLE OF DEGRADING EXTRACELLULAR MATRIX PROTEINS.
 CC - CASP DEGRADABLE TYPE I AND IV COLLAGEN AND FIBRONECTIN IN THE
 CC PRESENCE OF CALCIUM.
 CC - SUBUNIT: HETERODIMER, CONSISTING OF HEAVY AND LIGHT CHAINS WITH
 CC DISULFIDE BONDS. THE HEAVY CHAIN IS EXPECTED TO BE A REGULATORY
 CC SUBUNIT AND THE LIGHT CHAIN CONTAINS THE CATALYTIC SITE.
 CC - DOMAIN: THE GLU-RICH REGION IN THE N-TERMINAL REGION MAY BE GAMMA
 CC CARBOXYLATED AND FUNCTION AS A CALCIUM-BINDING SITE.
 CC - SIMILARITY: TO BLOOD COAGULATION FACTORS SUCH AS IX, X AND AN
 CC ANTICOAGULATION FACTOR, PROTEIN C.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC - SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC - SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>
 or send an email to license@lsb-sib.ch).

```

CC EMBL: X16160; CA34286.1; -
CC DR PIR: S05008; S05008.
CC DR HSSP: P00763; IDPO.
CC DR MEROPS: S01.193; -
CC DR InterPro: IPR000152; Asx_hydroxyl.
CC DR InterPro: IPR000859; CUB_domain.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR001881; EGF_Ca.
CC DR InterPro: IPR001254; Ser_protease_Try.
CC DR InterPro: IPR000436; Sushi_SCR_CCP.
CC DR Pfam: PF00008; EGF; 1.
CC DR Pfam: PF00084; sushi; 2.
CC DR Pfam: PF00089; trypsin; 1.
CC DR Pfam: PF00431; CUB; 2.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00033; CCP; 2.
CC DR SMART: SM00042; CUB; 2.
CC DR SMART: SM00179; EGF_CA; 1.
CC DR SMART: SM00020; Tryp_Spc; 1.
CC DR PROSITE: PS00010; ASX_HYDROXYL; 1.
CC DR PROSITE: PS01180; CUB; 2.

```


[illegible]

Query Match	29.0%	Score 69	DB 1	Length 986
Best Local Similarity	44.8%	Pred. NO. 1.4		
Matches	13	Conservative	4	Mismatches 12
				Indels 0
				Gaps 0
Qy	13	GRLASPGFPGETANDQERRRWTITAPGCR	41	
		: : : :		
Db	600	GI7TSPGWEKEYPNNKNCIMOLVAPTAQYR	638	

RESULT 14
BMP1_MOUSE

ID	NAME	STANDARD:	PRT:	991 AA.
AC	BMP1.MOUSE			
DT	P98063:			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bone morphogenetic protein 1 precursor (EC 3.4.24.19) (BMP-1)			
DE	(Procollagen C-proteinase) (PCP) (Mammalian tollold protein) (mTld).			
GN	BMP1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxId:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Embryo;			
RX	MEDLINE=94229342; PubMed=8174772;			
RA	Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;			
RT	"Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),			
RT	which is related to the drosophila dorsoventral gene tollold and			
RT	encodes a putative astacin metalloendopeptidase."			
RL	Dev. Biol. 153:175-183(1994).			
CC	-I- FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II			
CC	AND II. INDUCES CARTILAGE AND BONE FORMATION.			
CC	-I- CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at			
CC	Ala-I-Asp in type I and II procollagens and at Arg-I-Asp in type			
CC	III.			
CC	-I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	-I- ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-			
CC	ENDOPEPTIDASE ENHANCER PROTEIN.			
CC	-I- TISSUE SPECIFICITY: AT HIGH LEVELS IN EMBRYONIC MATERNAL DECIDUUM			
CC	AND FLOOR PLATE REGION OF THE NEURAL TUBE. LESS IN DEVELOPING			
CC	MEMBRANOUS AND ENDOCHONDRAL BONE, SUBMUCOSA OF INTESTINE, DERMIS			
CC	OF SKIN AND THE MESENCHYME OF SPLEEN AND LUNG.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.			
CC	-I- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.			
CC	-I- SIMILARITY: CONTAINS 5 CUB DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its use			
CC	by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: L24755; AAA37306.1; ..			
DR	HSSP: P00736; IAPQ.			
DR	MEROPS: M12.005; Bmp1.			
DR	MGI: 88176; Bmp1.			
DR	InterPro: IPR001506; Astacin.			
DR	InterPro: IPR000152; Asx_hydroxyl.			
DR	InterPro: IPR000859; CUB_domain.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR001881; EGF_Ca.			
DR	InterPro: IPR000130; Zn_MTPeptase.			
DR	Pfam: PF00008; EGF; 2.			
DR	Pfam: PF01400; Astacin; 1.			
DR	PRINTS: PR00480; ASPACTIN.			
DR	SMART: SMO0042; CUB; 5.			
DR	SMART: SMO0179; EGF_CA; 2.			
DR	SMART: SMO0235; ZNMG; 1.			
DR	PROSITE: PS00142; ZINC_PROTEASE; 1.			
DR	PROSITE: PS01180; CUB; 5.			
DR	PROSITE: PS00010; ASX_HYDROXYL; 2.			
DR	PROSITE: PS00022; EGF_1; FALSE_NEG.			
DR	PROSITE: PS01186; EGF_2; 2.			
DR	PROSITE: PS01187; EGF_CA; 2.			
KW	Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrolase;			
KW	Metalloproteinase; EGF-like domain; Zinc; Calcium; Signal;			
KW	Glycoprotein; Zymogen			
FT	SIGNAL	1	25	POTENTIAL.
FT	PROPER	26	125	POTENTIAL.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use.
CC

OC Eukaryota; Metazoa; Echinodermata;
OC Echinoida; Euechinoida; Echinacea
OC Echinacea

0C Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;
0C Strongylocentrotus.

OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=94215496; PubMed=8162855;
 RA Hwang S.P.L., Partin J.S., Lennarz W.J.,
 RT "Characterization of a homolog of human bone morphogenetic protein 1
 RL in the embryo of the sea urchin, Strongylocentrotus purpuratus.";
 CC Development 120:559-568(1994).
 CC -1- TISSUE SPECIFICITY: ECTODERMAL AND PRIMARY MESENCHYM CELLS IN
 CC HATCHED BLASTULA.
 CC -1- DEVELOPMENTAL STAGE: EMBRYO; HIGHEST LEVEL BEFORE SPICULOGENESIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L23838; AAA30081.1; -
 CC HSSP: P00736; IAPQ.
 CC MEROPS: M12.005; -
 CC InterPro: IPR001506; Astacin.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR000130; Zn_MTPetdase.
 CC Pfam: PF00008; EGF_1.
 CC Pfam: PF00431; CUB_2.
 CC Pfam: PF01400; Astacin; 1.
 CC PRINTS: PR00480; ASTACIN.
 CC SMART: SM00042; CUB; 2.
 CC SMART: SM00179; EGF_Ca; 1.
 CC SMART: SM00235; ZMNC; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS00010; ASX_HYDROXYL; 1.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS01187; EGF_Ca; 1.
 CC Developmental protein. Repeat; Hydrolase; Protease; Zinc; Embryo;
 CC Metalloprotease; EGF-like domain; Calcium; Signal.
 CC SIGNAL
 CC PROPEP 1 23
 CC CHAIN 110 639
 CC DOMAIN 110 306
 CC FT 307 419
 CC FT 420 531
 CC DOMAIN 532 573
 CC FT 197 197
 CC DOMAIN 198 198
 CC FT ACT_SITE 198 198
 CC FT METAL 201 201
 CC FT METAL 207 207
 CC FT METAL 536 548
 CC FT DISULFID 544 557
 CC FT DISULFID 559 572
 CC FT CARBOHYD 122 122
 CC FT CARBOHYD 140 140
 CC FT CARBOHYD 317 317
 CC FT CARBOHYD 455 455
 CC SEQUENCE 639 AA; 71893 MW; 59307B265B7894AD CRC64;
 Query Match 27.7%; Score 66; DB 1; Length 639;
 Best Local Similarity 40.7%; Pred. No. 2;
 Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

Db 316 GNFSPEMPCQYDQDCVWRISVTPG 342
 RESULT 17
 NRPI-CHICK
 ID NRPI-CHICK STANDARD; PRT; 914 AA.
 AC P79795;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein).
 GN NRPI OR NRP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 CC NCBI_TaxID=9031;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC SRRAIN-White Leghorn; Tissue=Embryonic brain;
 RX MEDLINE=95324761; PubMed=7601310;
 RA Takagi S., Kasuya Y., Shimizu M., Matsura T., Tsudoi M., Kawakami A.,
 RA Fujisawa H.;
 RT "Expression of a cell adhesion molecule, neuropilin, in the
 RT developing chick nervous system.";
 RL Dev. Biol. 170:207-222(1995).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
 CC PROPERTIES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
 CC (LAYERS D AND E OF SCES), AMACRINE CELLS OF RETINA, NEURITES OF
 CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
 CC BLOOD VESSELS IN THE ENTIRE EMBRYO.
 CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D45416; BA08256.1; -
 CC HSSP: P12259; IC2T.
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR000421; FAS8_C.
 CC InterPro: IPR000998; MAM_domain.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF00629; MAM; 1.
 CC Pfam: PF00754; F5_F8_Type_C; 2.
 CC PRINTS: PR00020; MAMDOMAIN.
 CC SMART: SM00042; CUB; 2.
 CC SMART: SM00231; FAS8C; 2.
 CC SMART: SM00137; MAM; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS01285; FAS8C_1; 2.
 CC PROSITE: PS01285; FAS8C_2; 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS00740; MAM_2; 1.
 CC Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Cell adhesion.
 FT SIGNAL 1 18
 FT CHAIN 19 914
 FT DOMAIN 20 847
 FT POTENTIAL.
 FT NEUROPILIN-1.
 FT EXTRACELLULAR (POTENTIAL).

```

CC TRANSMEM 848 870 POTENTIAL.
CC DOMAIN 871 914 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 25 139 CUB 1.
CC FT DOMAIN 145 263 CUB 2.
CC FT DOMAIN 273 422 F5/8 TYPE C 1.
CC FT DOMAIN 429 581 F5/8 TYPE C 2.
CC FT DOMAIN 636 801 MAM.
CC FT DISULFID 25 52 PROBABLE.
CC FT DISULFID 80 102 PROBABLE.
CC FT DISULFID 145 171 PROBABLE.
CC FT DISULFID 204 226 PROBABLE.
CC FT DISULFID 273 422 BY SIMILARITY.
CC FT DISULFID 429 581 BY SIMILARITY.
CC FT DISULFID 581 102480 MW; DD2E6D6FCBB68C CRC64;
CC SEQUENCE 914 AA: 102480 MW; DD2E6D6FCBB68C CRC64;

Query Match 27.7% Score 66; DB 1; Length 914;
Best Local Similarity 41.4%; Pred. No. 2.9;
Matches 12; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

13 GRASPGEPEYANDQERRMTLPAPGYR 41
| | | | | | | | | | | | | | | | | |
35 GYLTPSPQSYHPSQKCEWLQAPERYQ 63

RESULT 18
CA19_CHICK STANDARD; PRT; 503 AA.
AC P12106;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE collagen alpha 1(IX) chain precursor (Fragments).
GN COL9A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1;
RP SEQUENCE OF 1-416 FROM N.A.
RX MEDLINE=90062114; PubMed=2584206;
RA Nishimura I., Muregaki Y., Olsen B.R.;
RT Tissue-specific forms of type IX collagen-proteoglycan arise from
RL the use of two widely separated promoters."
RN J. Biol. Chem. 264:20033-20041(1989).
RN 12;
RP SEQUENCE OF 1-299 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88115376; PubMed=3339014;
RA Vasios G., Nishimura I., Konomi H., van der Rest M., Ninomiya Y.,
RT Olsen B.R.;
RT "Cartilage type IX collagen-proteoglycan contains a large
RT amino-terminal globular domain encoded by multiple exons."
RN J. Biol. Chem. 263:2324-2329(1988).
RN 3;
RP SEQUENCE OF 417-504 FROM N.A.
RX MEDLINE=85216631; PubMed=3858862;
RA Lozano G., Ninomiya Y., Thompson H., Olsen B.R.;
RT "A distinct class of vertebrate collagen genes encodes chicken type
RT IX collagen polypeptides."
RN Proc. Natl. Acad. Sci. U.S.A. 82:4050-4054(1985).
RN 1;
RP FUNCTION: Structural component of hyaline cartilage and vitreous
RN of the eye.
RN 1;
RP SUBUNIT: Heterotrimer of a alpha 1(IX), a alpha 2(IX) and a alpha
RN 3(IX) chain.
RN 1;
RP DOMAIN: EACH SUBUNIT IS COMPOSED OF THREE TRIPLE-HELICAL DOMAINS
RN INTERSPERSED WITH NONCOLLAGENOUS DOMAINS. THE GLOBULAR DOMAIN AT
RN THE N-TERMINUS OF TYPE IX COLLAGEN MOLECULES REPRESENTS THE NC4
RN DOMAIN WHICH MAY PARTICIPATE IN ELECTROSTATIC INTERACTIONS WITH
RN POLYANIONIC GLYCOSAMINOGLYCANS IN CARTILAGE.
RN 1;
RP PTM: Covalently linked to the telopeptides of type II collagen by
RN lysine-derived cross-links.
RN 1;
RP PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
RN UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

```

```

CC 1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M28659; AAA48708.1; -
CC EMBL; J03539; AAA48702.1; -
CC EMBL; M11325; AAA48642.1; -
CC EMBL; M11324; AAA48642.1; JOINED.
CC PIR; A28360; A28360.
CC PIR; A28754; A28754.
CC PIR; B34493; B34493.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR003129; TSPN.
CC SMART; SM00210; TSPN; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Cartilage; Signal.
CC SIGNAL 1 23
CC CHAIN 24 >416 COLLAGEN ALPHA 1(IX) CHAIN.
CC FT NON CONS 416 417
CC FT CHAIN 24 >416 COLLAGEN ALPHA 1(IX) CHAIN.
CC FT CHAIN 417 503
CC FT DOMAIN 24 266 NONHELICAL REGION (N-TERMINAL) (NC4).
CC FT DISULFID 44 198 POTENTIAL.
CC FT DISULFID 242 252 POTENTIAL.
CC FT CONFLICT 25 25 Y -> I (IN REF. 2).
CC FT CONFLICT 51 51 E -> Q (IN REF. 2).
CC SEQUENCE 503 AA; 51879 MW; DDC0808CDC70A7C5 CRC64;

Query Match 27.5% Score 65.5; DB 1; Length 503;
Best Local Similarity 42.1%; Pred. No. 1.8;
Matches 16; Conservative 2; Mismatches 15; Indels 5; Gaps 1;

QY 2 PLGPKPEPYFGRGLASPGGEYANDQERRMTLPAPG 39
| | | | | | | | | | | | | | | | | |
DB 298 PPGPGPDGAGKAGSPGLGEPGAD-----GLTGPDG 330

RESULT 19
TLD_DROME STANDARD; PRT; 1057 AA.
AC P25723; Q9VC46;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dorsal-ventral patterning tolloid protein precursor (EC 3.4.24.-).
GN TLD OR CG6868.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1;
RP SEQUENCE FROM N.A.
RN STRAIN=Cancon-S;
RP MEDLINE=92034970; PubMed=1840509;
RA Shinnell M.J., Ferguson E.L., Childs S.R., O'Connor M.B.;
RT "The Drosophila dorsal-ventral patterning gene tolloid is related to
RT human bone morphogenetic protein 1."
RN Cell 67:469-481(1991).
RN 12;
RP SEQUENCE FROM N.A.
RX MEDLINE=95324373; PubMed=7600963;
RA Finelli A.L., Bossie C.A., Xie T., Padgett R.W.;
RT "Mutational analysis of the Drosophila tolloid gene, a human BMP-1
RT homolog."
RN Development 120:861-870(1994).

```


DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-Sprague-Dawley;
RX	MEDLINE=97433085; PubMed=928754;
RA	Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,
RT	Ginty D.D.;
RL	"Neuropilin is a semaphorin III receptor.";
CL	Cell 90:753-762(1997).
CC	- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC	AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE
CC	CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE
CC	LINING IN THE RIBS.
CC	- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC	- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF016297; AAC53338.1; -
DR	HSSP; P12259; ICGT.
DR	InterPro; IPR000859; CUB_domain.
DR	InterPro; IPR000421; FA58_C.
DR	InterPro; IPR000998; MAM_domain.
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00629; MAM; 1.
DR	Pfam; PF00754; F5_F8_Type_C; 2.
DR	SMART; SM00042; CUB; 2.
DR	SMART; SM00231; FA58C; 2.
DR	SMART; SM00137; MAM; 1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PROSITE; PS50060; MAM_2; 1.
DR	Transmembrane; Glycoprotein; Neutrine; Signal; Repeat; Receptor.
FT	SIGNAL 1 22
FT	CHAIN 23 925
FT	DOMAIN 23 858
FT	TRANSMEM 859 883
FT	DOMAIN 884 925
FT	DOMAIN 28 142
FT	DOMAIN 149 267
FT	DOMAIN 277 427
FT	DOMAIN 434 592
FT	DOMAIN 642 802
FT	DOMAIN 28 55
FT	DISULFID 83 105
FT	DISULFID 149 175
FT	DISULFID 208 230
FT	DISULFID 277 427
FT	DISULFID 434 592
FT	CARBOHYD 152 152
FT	CARBOHYD 157 157
FT	CARBOHYD 629 629
FT	CARBOHYD 833 833
FT	CARBOHYD 834 834
FT	SEQUENCE 925 AA: 103896 MW: 38629035F644851C CRG4;

Query Match	26.9%	Score 64	DB 1	Length 925
Best Local Similarity	33.3%	Pred. No. 5.1		
Matches	12	Conservative	6	Mismatches 18; Indels 0; Gaps 0

QY	2	P	L	G	R	K	M	P	E	P	V	R	G	R	L	A	S	P	E	F	C	E	A	D	O	E	R	R	M	L	T	A	P	37			
DB	27	P	C	G	G	R	L	N	S	K	D	M	A	G	T	T	S	P	G	I	P	O	D	I	P	S	H	O	N	C	E	W	V	I	A	P	62

Search completed: January 11, 2003, 10:49:28
Job time : 67 secs

PD 24-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
PR 01-JUN-2001; 2001DK-0000870.
XX
XX (JENS/) JENSENIUS J C.
PA (THIE/) THIEL S.
XX
XX Jansenius JC, Thiel S;
PI WPI; 2002-179791/23.
XX
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX
XX Claim 39; Page 3; 76pp; English.
XX
XX The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human MASP-2 20 kDa fragment.
XX
SQ Sequence 41 AA;
XX
XX Query Match 100.0%; Score 238; DB 23; Length 41;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TPLGPKWPEVFGRLASPGFGEYANDQERRRWLTAPPGYR 41
Db 1 TPLGPKWPEVFGRLASPGFGEYANDQERRRWLTAPPGYR 41
XX
XX 17-MAY-2002 (first entry)
XX
XX Human mature MASP-2 protein.
XX
XX Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW Complement fixation; Infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Region 15..671
FT /note= "This region is specifically claimed in
FT claim 45"
XX
XX WO200206460-A2.
XX
XX 24-JAN-2002.
XX

PF 13-JUL-2001; 2001WO-DK00499.
XX
XX 13-JUL-2000; 2000DK-0001089.
PR 01-JUN-2001; 2001DK-0000870.
XX
XX (JENS/) JENSENIUS J C.
PA (THIE/) THIEL S.
XX
XX Jansenius JC, Thiel S;
PI WPI; 2002-179791/23.
XX
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX
XX Claim 45; Page 73-76; 76pp; English.
XX
XX The invention relates to use of a polypeptide derived from
CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
CC producing a pharmaceutical composition. MASP-2 is a complement-
CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medicament or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is human mature MASP-2 protein.
XX
SQ Sequence 671 AA;
XX
XX Query Match 100.0%; Score 238; DB 23; Length 671;
Best Local Similarity 100.0%; Pred. No. 5.4e-22;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TPLGPKWPEVFGRLASPGFGEYANDQERRRWLTAPPGYR 41
Db 1 TPLGPKWPEVFGRLASPGFGEYANDQERRRWLTAPPGYR 41
XX
XX RESULT 3
XX ID AAE14564 standard; Protein; 686 AA.
XX
XX AAE14564;
XX
XX 17-MAY-2002 (first entry)
XX
XX Human MASP-2 protein.
XX
XX Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
KW Complement fixation; Infection; microbe; retrovirus; HIV; abortion;
KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Peptide 1..15
FT /label= Signal_peptide
FT Protein 16..686
FT /label= Mature_MASP-2_protein
FT Region 16..56
FT /note= "MASP-2 20 kda fragment"
FT Domain 16..31
FT /label= C1r/C1s-like_domain
FT Region 16..296
FT /note= "This region is specifically claimed in claim 46"
FT

FT Region 30..444
/note= "This region is specifically claimed in claim 41."
FT The region 30 to 296 is specifically claimed in claim 47"
FT 108..134
FT /note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT 135..146
FT /label= EGF-like_domain
FT Region 138..296
/note= "This region is specifically claimed in claim 43"
FT Misc-difference 155
/note= "Encoded by CA"
FT Misc-difference 156
/note= "Encoded by C"
FT Domain 183..199
/label= C1r/C1s-like_domain
FT Domain 293..307
/label= CCP-1_domain
/note= "Complement control protein domain"
FT Domain 363..376
/label= CCP-2_domain
FT Region 377..388
/note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT Region 410..417
/note= "Fragment obtained by amino acid sequencing
FT of peptides"
FT Region 432..441
/note= "Linker"
FT Domain 445..468
/label= Serine_protease_domain
FT Active-site 483
FT Active-site 532
FT Active-site 633
XX WO200206460-A2.
XX PD 24-JAN-2002.
XX 13-JUL-2001; 2001WO-DK00499.
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000870.
XX PA (JENS/) JENSENIUS J C.
PA (THIE/) THIEL S.
XX Jensenius JC, Thiel S;
XX MPI: 2002-179791/23.
XX N-PSDB: AAD24224.
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX Claim 41: Fig 6; 76pp: English.
XX The invention relates to use of a polypeptide derived from
XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
XX producing a pharmaceutical composition. MASP-2 is a complement-
XX fixing enzyme and involved in lectin pathway of complement activation.
XX The pharmaceutical composition comprising MASP-2 is useful for
XX treating infections caused by microbes such as fungus, yeast,
XX retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
XX bacteria which are resistant to at least one antibiotic medicament or
XX multiresistant. The polynucleotide encoding MASP-2 is useful for
XX treating patients deficient in MASP-2. The invention also discloses
XX MASP-2 assays which are useful for determination of MASP-2 activity or
XX levels in patients suffering from e.g. infections, inflammatory disorders
XX and spontaneous recurrent abortion. The pharmaceutical composition
XX comprising MASP-2 inhibitor is useful for treating inflammatory

CC disorders. The present sequence is human MASP-2 protein.
CC Note: The present sequence is stated as being the same as
CC SEQ ID NO:2 shown in sequence listing of the specification (AAE14568).
CC However the sequences differ at various locations.
XX
XX SQ Sequence 686 AA:
XX
XX Query Match 100.0%; Score 238; DB 23; Length 686;
XX Best Local Similarity 100.0%; Pred. No.5.5e-22;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TPGLPKWPEPVFGRLASPGFPGEXANDQERRMTLTAPGGR 41
XX 16 TPGLPKWPEPVFGRLASPGFPGEXANDQERRMTLTAPGGR 56
XX
XX RESULT 4
XX ID AAE14568 standard; Protein; 686 AA.
XX AC AAE14568;
XX DT 17-MAY-2002 (first entry)
XX DE Human MASP-2 protein, alternative version.
XX XX
XX KW Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL;
XX KW complement fixation; infection; microbe; retrovirus; HIV; abortion;
XX KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..15
XX FT Protein /label= Signal_peptide
XX FT 16..686
XX FT /label= Mature_MASP-2_protein
XX FT Misc-difference 155
XX FT /note= "Encoded by CA"
XX FT Misc-difference 156
XX FT /note= "Encoded by C"
XX FT Misc-difference 298
XX FT /note= "Encoded by CAG"
XX FT Misc-difference 299
XX FT /note= "Encoded by CCF"
XX
XX WO200206460-A2.
XX PN 24-JAN-2002.
XX 13-JUL-2001; 2001WO-DK00499.
XX 13-JUL-2000; 2000DK-0001089.
XX 01-JUN-2001; 2001DK-0000870.
XX PA (JENS/) JENSENIUS J C.
PA (THIE/) THIEL S.
XX Jensenius JC, Thiel S;
XX PI MPI: 2002-179791/23.
XX DR N-PSDB: AAD24224.
XX Use of a polypeptide comprising amino acid sequence derived from
PT mannan-binding lectin associated serine protease-2 (MASP-2) for
PT producing pharmaceutical composition, to treat bacterial, fungal, viral
PT infections -
XX Claim 41: Page 71-73; 76pp: English.
XX The invention relates to use of a polypeptide derived from
XX mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
XX producing a pharmaceutical composition. MASP-2 is a complement-

CC fixing enzyme and involved in lectin pathway of complement activation.
CC The pharmaceutical composition comprising MASP-2 is useful for
CC treating infections caused by microbes such as fungus, yeast,
CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
CC bacteria which are resistant to at least one antibiotic medication or
CC multiresistant. The polynucleotide encoding MASP-2 is useful for
CC treating patients deficient in MASP-2. The invention also discloses
CC MASP-2 assays which are useful for determination of MASP-2 activity or
CC levels in patients suffering from e.g. infections, inflammatory disorders
CC and spontaneous recurrent abortion. The pharmaceutical composition
CC comprising MASP-2 inhibitor is useful for treating inflammatory
CC disorders. The present sequence is stated as being the same as
CC Note: The present sequence is stated as being the same as
CC SEQ ID NO:2 shown in figure 6 of the specification (AAB14564).
CC However the sequences differ at various locations.
XX
SQ Sequence 686 AA:
Query Match 100.0%; Score 238; DB 23; Length 686;
Best Local Similarity 100.0%; Pred. No. 5.5e-22;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPPGYR 41
|||||
16 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPPGYR 56
Db.
RESULT 5
ABG21134
ID ABG21134 standard; Protein; 1034 AA.
XX
AC ABG21134;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21125.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO2001/5067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PP 31-MAR-2000; 2000US-0540217. } the lute
PP 23-AUG-2000; 2000US-0649167. }
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS85321.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID NO 51493; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1034 AA:
Query Match 100.0%; Score 238; DB 22; Length 1034;
Best Local Similarity 100.0%; Pred. No. 8.7e-22;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPPGYR 41
|||||
Db 752 TPLGPKWPEPVFGRLASPGPGGYANDQERRWTLTAPPGYR 792
XX
RESULT 6
AAU87257
ID AAU87257 standard; Protein; 174 AA.
XX
AC AAU87257;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #167.
XX
XX
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200155318-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
PP 17-JAN-2001; 2001WO-US01332.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.

CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 39.3%; Score 93.5; DB 22; Length 174;
Best Local Similarity 40.4%; Pred. No. 0.00039;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

OY 5 PKW--DEPV-----FGRLASPGFPGEVANDQERWTLTPGQYR 41
DB 69 PXWDSKEPVCIACGGVIRNATGTGRVSPGFGNYSNLTCHMLLEAPESQR 120

RESULT 7

ABBI0255 standard; Protein; 174 AA.

ABBI0255;

10-JAN-2002 (first entry)

Human cDNA SEQ ID NO: 563.

Human; gene therapy; neural disorder; immune system disorder;
muscular disorder; reproductive disorder; gastrointestinal disorder;
pulmonary disorder; cardiovascular disorder; renal disorder;
proliferative disorder; inflammation.

Homo sapiens.

WO200154474-A2.

02-AUG-2001.

17-JAN-2001; 2001MO-US01349.

31-JAN-2000; 2000US-179065P.
04-FEB-2000; 2000US-180628P.
24-FEB-2000; 2000US-184664P.
02-MAR-2000; 2000US-186350P.
16-MAR-2000; 2000US-189874P.
17-MAR-2000; 2000US-190076P.
18-APR-2000; 2000US-198133P.
19-MAY-2000; 2000US-205515P.
27-JUN-2000; 2000US-209467P.
28-JUN-2000; 2000US-214886P.
30-JUN-2000; 2000US-215135P.
07-JUL-2000; 2000US-216647P.
07-JUL-2000; 2000US-216880P.
11-JUL-2000; 2000US-217487P.
11-JUL-2000; 2000US-217496P.
14-JUL-2000; 2000US-218290P.
26-JUL-2000; 2000US-220963P.
26-JUL-2000; 2000US-220964P.
14-AUG-2000; 2000US-224518P.
14-AUG-2000; 2000US-224519P.
14-AUG-2000; 2000US-225213P.
14-AUG-2000; 2000US-225214P.
14-AUG-2000; 2000US-225266P.
14-AUG-2000; 2000US-225267P.
14-AUG-2000; 2000US-225268P.
14-AUG-2000; 2000US-225270P.
14-AUG-2000; 2000US-225447P.
14-AUG-2000; 2000US-225757P.
14-AUG-2000; 2000US-225758P.
14-AUG-2000; 2000US-225759P.
18-AUG-2000; 2000US-226279P.
22-AUG-2000; 2000US-226681P.

22-AUG-2000; 2000US-226688P.
22-AUG-2000; 2000US-227182P.
23-AUG-2000; 2000US-227009P.
30-AUG-2000; 2000US-228924P.
01-SEP-2000; 2000US-229287P.
01-SEP-2000; 2000US-229343P.
01-SEP-2000; 2000US-229344P.
01-SEP-2000; 2000US-229345P.
05-SEP-2000; 2000US-229509P.
05-SEP-2000; 2000US-229513P.
06-SEP-2000; 2000US-230437P.
06-SEP-2000; 2000US-230438P.
08-SEP-2000; 2000US-231242P.
08-SEP-2000; 2000US-231243P.
08-SEP-2000; 2000US-231244P.
08-SEP-2000; 2000US-231413P.
08-SEP-2000; 2000US-231414P.
08-SEP-2000; 2000US-232080P.
08-SEP-2000; 2000US-232081P.
12-SEP-2000; 2000US-231968P.
14-SEP-2000; 2000US-232397P.
14-SEP-2000; 2000US-232398P.
14-SEP-2000; 2000US-232399P.
14-SEP-2000; 2000US-232400P.
14-SEP-2000; 2000US-232401P.
14-SEP-2000; 2000US-233063P.
14-SEP-2000; 2000US-233064P.
14-SEP-2000; 2000US-233065P.
21-SEP-2000; 2000US-234223P.
21-SEP-2000; 2000US-234274P.
25-SEP-2000; 2000US-234998P.
25-SEP-2000; 2000US-235484P.
26-SEP-2000; 2000US-235834P.
27-SEP-2000; 2000US-235836P.
29-SEP-2000; 2000US-236327P.
29-SEP-2000; 2000US-236367P.
29-SEP-2000; 2000US-236368P.
29-SEP-2000; 2000US-236369P.
29-SEP-2000; 2000US-236370P.
02-OCT-2000; 2000US-236802P.
02-OCT-2000; 2000US-237037P.
02-OCT-2000; 2000US-237038P.
02-OCT-2000; 2000US-237039P.
02-OCT-2000; 2000US-237040P.
13-OCT-2000; 2000US-239935P.
13-OCT-2000; 2000US-239937P.
20-OCT-2000; 2000US-240960P.
20-OCT-2000; 2000US-241221P.
20-OCT-2000; 2000US-241785P.
20-OCT-2000; 2000US-241786P.
20-OCT-2000; 2000US-241787P.
20-OCT-2000; 2000US-241808P.
20-OCT-2000; 2000US-241809P.
20-OCT-2000; 2000US-241826P.
01-NOV-2000; 2000US-244617P.
08-NOV-2000; 2000US-246474P.
08-NOV-2000; 2000US-246475P.
08-NOV-2000; 2000US-246476P.
08-NOV-2000; 2000US-246477P.
08-NOV-2000; 2000US-246478P.
08-NOV-2000; 2000US-246523P.
08-NOV-2000; 2000US-246524P.
08-NOV-2000; 2000US-246525P.
08-NOV-2000; 2000US-246526P.
08-NOV-2000; 2000US-246527P.
08-NOV-2000; 2000US-246528P.
08-NOV-2000; 2000US-246532P.
08-NOV-2000; 2000US-246609P.
08-NOV-2000; 2000US-246610P.
08-NOV-2000; 2000US-246611P.
08-NOV-2000; 2000US-246613P.
17-NOV-2000; 2000US-249207P.

PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234988.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX XX
XX XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
DR WPI: 2001-465568/50.
DR N-PSDB; AAS31599.
XX
XX
PT Isolated nucleic acid molecule encoding a calcium-binding protein is
PS used in preventing, treating or ameliorating a medical condition -
PS Claim 11; SEQ ID No 111; 542pp; English.
XX
XX The present invention relates to the isolation of novel human
CC calcium-binding proteins, and cDNA (AAS31577-AAS31654) and genomic
CC sequences encoding for these proteins. The sequences of the invention
CC are useful in the diagnosis, prevention and/or prognosis of diseases
CC associated with aberrant calcium flux. Such disorders include
CC neurological diseases (e.g. amyotrophic lateral sclerosis, ALS),
CC immune dysfunction (e.g. severe combined immunodeficiency, SCID),
CC digestive disorders (e.g. irritable bowel syndrome, IBS), neoplastic
CC disease (e.g. cancer), blood disorders (e.g. haemophilia), and/or
CC infectious disease (e.g. acquired immunodeficiency syndrome, AIDS). The
CC novel calcium-binding proteins are also useful as screening tools to
CC identify antagonists and/or agonists that may enhance or inhibit
CC activities mediated by calcium-binding proteins. The polynucleotides of
CC the invention are also useful in gene therapy. AAU19892-AAU19969
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 174 AA;
Query Match 39.3%; Score 93.5; DB 22; Length 174;
Best Local Similarity 40.4%; Pred. No. 0.00039;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;
QY 5 PKW-PEPV-----FGRLASPGFGEVANDQERRWTLTAPPQYR 41
DB 69 PXMWSEPVCIACGGVIRNATGRIVSPGPNYSNNLTCWHLBAPQGR 120
RESULT 9
AAG01729 ID AAG01729 standard; Protein; 60 AA.
XX
AC AAG01729;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5810.
XX
XX Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX

DR		MP1: 2000-500381/45.
DR	N-PSDB; AAC01735.	
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
XX		
PS	Claim 13: SEQ ID 5810; 71pp + CD-ROM; English.	
CC		
XX	The present sequence is a polypeptide encoded by one of a large number	
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs	
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30'	
CC	different tissues. EST sequences usually correspond mainly to the 3'	
CC	untranslated region (UTR) of the mRNA because they are often obtained	
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for	
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in	
CC	those cases where longer cDNA sequences have been obtained, the full 5'	
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'	
CC	ends and can therefore be used to obtain full length cDNAs and genomic	
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and	
CC	chromosome mapping procedures. They are used to obtain upstream	
CC	regulatory sequences and to design expression and secretion vectors.	
SO	Sequence 60 AA:	
Query Match	39.1%; Score 93; DB 21; Length 60;	
Best Local Similarity	45.2%; Pred. No. 0.00014;	
Matches 14: Conservative 7; Mismatches 10; Indels 0; Gaps 0		
Oy	11 VFGRLASPGFGEVANDDERRWTLTAAPGGR 41	
.db	: : : : : : : : : : : : : : : : :	
	1 MFGQIOSPEGYPSDSEVTWNTVPDGFR 31	
RESULT 10		
AAG00221		
ID	AAG00221 standard; Protein: 97 AA.	
XX		
AC	AAG00221;	
XX		
DT	06-OCT-2000 (first entry)	
XX		
DE	Human secreted protein, SEQ ID NO: 4302.	
XX		
KW	Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.	
XX		
XX	omo sapiens.	
XX	EP1033401-A2.	
PD	06-SEP-2000.	
XX		
PF	21-FEB-2000; 2000EP-0200610.	
XX		
PR	26-FEB-1999; 99US-0122487.	
XX		
PA	(GEST) GENSET.	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX		
DR	WP1: 2000-500381/45.	
DR	N-PSDB; AAC00227.	
XX		
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
XX		
PS	Claim 13: SEQ ID 4302; 71pp + CD-ROM; English.	
CC		
CC	The present sequence is a polypeptide encoded by one of a large number	
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs	
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30'	
CC	different tissues. EST sequences usually correspond mainly to the 3'	
CC	untranslated region (UTR) of the mRNA because they are often obtained	
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for	
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in	
CC	those cases where longer cDNA sequences have been obtained, the full 5'	
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'	
CC	ends and can therefore be used to obtain full length cDNAs and genomic	
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and	
CC	chromosome mapping procedures. They are used to obtain upstream	
CC	regulatory sequences and to design expression and secretion vectors.	

CC		different tissues. EST sequences usually correspond mainly to the 3'
CC		untranscribed region (UTR) of the mRNA because they are often obtained
CC		from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC		isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC		those cases where longer cDNA sequences have been obtained, the full 5'
CC		UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC		ends and can therefore be used to obtain full length cDNAs and genomic
CC		DNA's. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC		chromosome mapping procedures. They are used to obtain upstream
CC		regulatory sequences and to design expression and secretion vectors.
XX		
SX	Sequence	97 AA:
.	Query Match	39.1%; Score 93; DB 21; Length 97;
.	Best Local Similarity	45.2%; Pred. NO. 0.00024;
Matches	14: Conservative	7: Mismatches 10: Indels 0: Gaps 0:
OY	11 VEGRLASPGFGEFYANDERRMTLTAPPCYR 41	
	: : : : : : : :	
Db	27 MFGQIQSPGYPSDYSSEVTWNTVPOGFR 57	
RESULT 11		
AAB47559		
ID	AAB47559 standard; Protein; 728 AA.	
AC		
XX	AAB47559;	
DT	13-DEC-2001 (first entry)	
XX		
XX	Protease PRTS-1.	
DE		
KM	Human; protease; PRTS; gastrointestinal disorder; peptic esophagitis;	
KM	indigestion; gastritis; cardiovascular disorder; reproductive disorder;	
KM	hypertensive heart disease; myocardial infarction; autoimmune disorder;	
KM	inflammatory disorder; rheumatoid arthritis; cell proliferative disorder;	
KM	arteriosclerosis; cancer; epithelial disorder; eczema; endometriosis;	
KM	neurological disorder; Huntington's disease.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..19
FT	/label=	Potential signal peptide
FT	Modified-site	42
FT	/label=	Potential phosphorylation site
FT	Modified-site	49
FT	/label=	Potential glycosylation site
FT	Modified-site	51
FT	/label=	Potential phosphorylation site
FT	Modified-site	96
FT	/label=	Potential phosphorylation site
FT	Modified-site	117
FT	/label=	Potential phosphorylation site
FT	Modified-site	123
FT	/label=	Potential phosphorylation site
FT	Modified-site	129
FT	/label=	Potential phosphorylation site
FT	Modified-site	167
FT	/label=	Potential phosphorylation site
FT	Modified-site	178
FT	/label=	Potential phosphorylation site
FT	Modified-site	180
FT	/label=	Potential phosphorylation site
FT	Modified-site	191
FT	/label=	Potential phosphorylation site
FT	Modified-site	198
FT	/label=	Potential phosphorylation site
FT	Modified-site	214
FT	/label=	Potential phosphorylation site
FT	Modified-site	293
FT	/label=	Potential phosphorylation site
FT	Modified-site	356
FT	/label=	Potential phosphorylation site

FT Modified-site /label= Potential phosphorylation site
FT 361
FT /label= Potential phosphorylation site
FT 381
FT Modified-site /label= Potential phosphorylation site
FT 385
FT Modified-site /label= Potential glycosylation site
FT 388
FT Modified-site /label= Potential phosphorylation site
FT 396
FT Modified-site /label= Potential phosphorylation site
FT 407
FT Modified-site /label= Potential glycosylation site
FT 472
FT Modified-site /label= Potential phosphorylation site
FT 476..531
FT Active-site /label= Serine protease active site
FT 501
FT Modified-site /label= Potential phosphorylation site
FT 506
FT Modified-site /label= Potential phosphorylation site
FT 533
FT Modified-site /label= Potential glycosylation site
FT 599
FT Modified-site /label= Potential glycosylation site
FT 601
FT Modified-site /label= Potential phosphorylation site
FT 612
FT Modified-site /label= Potential phosphorylation site
FT 619
FT Modified-site /label= Potential phosphorylation site
FT 632
FT Modified-site /label= Potential phosphorylation site
FT 640
FT Modified-site /label= Potential glycosylation site
FT 642
FT Modified-site /label= Potential phosphorylation site
FT 643..692
FT Active-site /label= Potential phosphorylation site
FT 675
FT Modified-site /label= Serine protease active site
FT 720
FT Modified-site /label= Potential phosphorylation site
FT /label= Potential phosphorylation site

XX WO200171004:A2.
XX
XX
XX 27-SEP-2001.
XX
XX
XX 16-MAR-2001; 2001MO-US08441.
XX
XX 17-MAR-2000; 2000US-190708P.
XX 30-MAR-2000; 2000US-193182P.
XX 14-APR-2000; 2000US-197086P.
XX 20-APR-2000; 2000US-199022P.
XX 28-APR-2000; 2000US-200227P.
XX
XX
XX (INCY-) INCYTE GENOMICS INC.

PI Yue H, Lu DAM, Policky JL, Delegeane AM, Tribouley CM, Khan FA;
PI Au-Young J, Bandman O, Lal P, Borowsky ML, Gandhi AR, Hillman JL;
PI Tang YT, Burrford N, Baughn MK, Nguyen DB, Yao MG, Walla NK, He A,
PI Hafalla A, Lu Y, Patterson C;
XX
XX WPI; 2001-611509/70.
DR N-PSDB; AAAH43512.
XX
XX New polypeptide for treating gastrointestinal, cardiovascular and
PT autoimmune disorders, comprises novel human proteases (PRTS) and
PT polynucleotides -
XX
XX Claim 1; Page 109-11; 129pp: English.
XX
XC The sequences given in AAH47559-69 show novel human proteases PRTS-1

. CC to -11. These proteases can be administered in pharmaceutical compounds
CC for the treatment of diseases or conditions associated with decreased
CC expression of functional protease. Disorders which can be treated
CC include gastrointestinal (peptic esophagitis, indigestion and
CC gastritis), cardiovascular (hypertensive heart disease, myocardial
CC infarction), autoimmune/inflammatory (rheumatoid arthritis), cell
CC proliferative (arteriosclerosis, cancers), epithelial (eczema),
CC neurological (Huntington's disease), and reproductive (endometriosis)
CC disorders. This protein shows homology with the human precursor of
CC p100 serine protease of Ra-reactive factor.

XX Sequence 728 AA;

SQ Query Match 39.1%; Score 93; DB 22; Length 728;
Best Local Similarity 45.2% ; Pred. No. 0.0021;
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0.

OY 11 VFGRASPGEVANNDOERMTLAPGYR 41
Db 27 MFQIQQPEYPDPSDEVTWNTVPDSFR 57

RESULT 12
AAB85060 AAB85060 standard; Protein; 728 AA.
AC AAB85060;
AD
AE
AF 06-AUG-2001 (first entry)
AG
AH Human serine protease MASP-3 polypeptide.
AI
AJ
AK Mannan-binding lectin associated serine protease-3; MASP-3; MBL; human;
AL mannan-binding lectin complement infection; cancer; cytostatic;
AM immunomodulator; neuroprotective; anticonvulsive; anti-inflammatory;
AN antiarthritic; anti-neuritic; anti-infectious; antipsoriatic; antidiabetic;
AO antithrombotic.
AP
AQ Homo sapiens.
AR
AS WO2001140451-A2.
AT
AU
AV 07-JUN-2001.
AW
AX 30-NOV-2000; 2000MO-DK00659.
AY PF
AZ 02-DEC-1999; 99DK-0001721.
BA PR
BB 21-JUL-2000; 2000DK-0001126.
BC
BD XX
BE (JENS//) JENSENIUS J C.
BF PA
BG (THIEU/) THIEL S.
BH
BI Jensenius JC, Thiel S:
BJ
BK WPI: 2001-374820/39,
BL DR N-PSDB; AAF83895.
BM XX
BN Novel pure mannan-binding lectin associated serine protease
BO polypeptides and polynucleotides encoding the protein, used for
BP treating reoxygenated ischemic tissues, mannon-binding lectin
BQ deficiency, multiple sclerosis -
BR Claim 15; Page 88-96; 99pp; English.
BS
BT
BU
BV The invention relates to a mannan-binding lectin (MBL) associated serine
BX protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
BY the complement activation particularly when bound to MBL/MASP-2 complexes
CZ and directly activates complement system through binding to MBL. The
CA MASP-3 polypeptides, polynucleotides and modulators are useful for
CB preparing a pharmaceutical composition for treating aberrant MASP-3
CD activity such as infections, cancer, MBL-deficiency, disorders of the
CE immune system and reproductive system. The MASP-3 polypeptide is also
CF immune system and reproductive system. The MASP-3 polypeptide is also
CG used for treating diseases associated with human immunodeficiency virus,

DB 403 PFWDSEPCVACGCGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

RESULT 15
ID AAB70541 standard; Protein: 525 AA.
XX
AC AAB70541;
XX
XX
DT 09-MAY-2001 (first entry)
XX
DE Human PRO11 protein sequence SEQ ID NO:22.
XX
KW Human; PRO; PROX; cytosolic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia.
OS Homo sapiens.
XX
XX WO200110902-A2.
XX
XX -5-FEB-2001.
XX
PF 11-AUG-2000; 2000MO-US21857.
XX
XX -11-AUG-1999; 99US-0148433.
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
DR N-PSDB; AAF74442.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
XX
PS Claim 1; Page 38-40; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytosolic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 525 AA;

Query Match 38.4%; Score 91.5; DB 22; Length 525;
Best Local Similarity 40.4%; Pred. No. 0.0023;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

QY 5 PKW--PEPV-----FGRLASGPFGEYANDQERRMTLTAPPQYR 41
DB 403 PFWDSEPCVACGCGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

RESULT 16
ID AAB70542 standard; Protein: 526 AA.
XX
AC AAB70542;
XX
XX
DT 09-MAY-2001 (first entry)

XX
DE Human PRO12 protein sequence SEQ ID NO:24.
XX
KW Human; PRO; PROX; cytosolic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia.
OS Homo sapiens.
XX
XX WO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX 11-AUG-2000; 2000MO-US21857.
PF 11-AUG-1999; 99US-0148433.
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
DR N-PSDB; AAF74443.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
XX
PS Claim 1; Page 41-43; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytosolic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 526 AA;

Query Match 38.4%; Score 91.5; DB 22; Length 526;
Best Local Similarity 40.4%; Pred. No. 0.0023;
Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2;

QY 5 PKW--PEPV-----FGRLASGPFGEYANDQERRMTLTAPPQYR 41
DB 403 PFWDSEPCVACGCGVIRNGTGRIVSPGPGNYSNLTCHMLLEAPESQR 454

RESULT 17
ID AAE15854 standard; Protein: 829 AA.
XX
AC AAE15854;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Human SEZ6 mature protein.
XX
XX Human; SEZ6; neural regeneration; seizure; infertility; gene therapy;
KW stroke; Alzheimer's disease; Huntington's disease; myasthenia gravis;
KW neurotrophic; neuroprotective; anticonvulsant; cerebroprotective; stroke;
KW vasotropic; neuronal growth; growth factor-mediated chemotaxis; trauma;
KW neurological disorder; muscular dystrophy; muscle injury; vulnery;
KW amyotrophic lateral sclerosis; multiple sclerosis; ischemia; diabetes;
KW epilepsy; Parkinson's disease; sexual development; gene mapping;


```

AAB70537
ID AAB70537 standard; Protein: 993 AA.
AC AAB70537;
XX
XX 09-MAY-2001 (first entry)
XX
XX Human PRO7 protein sequence SEQ ID NO:14.
XX
XX Homo sapiens.
XX
XX WO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX 11-AUG-2000; 2000WO-US21857.
XX
XX 11-AUG-1999; 99US-0148433.
XX 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
XX
XX N-PSDB: AAF74438.
XX
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
XX polypeptides, useful for treating a syndrome associated with a
XX PROX-associated disorder, e.g. cancer -
XX
XX Claim 1; Page 24-27; 16pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
XX polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
XX X is an integer from 1 to 17). PROX polypeptides have cytostatic,
XX immunomodulatory and reproduction activities, and can be used in gene
XX therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
XX nucleic acids and antibodies are useful in the manufacture of a
XX medicament for treating a syndrome associated with a PROX-associated
XX disorder, e.g. a cell proliferation and/or differentiation disorder
XX (e.g. cancer or immune associated disorders) and a gestational disease
XX (e.g. pre-clampsia). They are also used for screening for a modulator of
XX activity or of latency or predisposition to a PROX-associated disorder.
XX AAF74432 to AAF7448 encode the specifically claimed human PROX
XX polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
XX Sequence 993 AA:
XX
XX Query Match 38.4%; Score 91.5; DB 22; Length 993;
XX Best Local Similarity 40.4%; Pred. No. 0.0046;
XX Matches 21; Conservative 3; Mismatches 13; Indels 15; Gaps 2.
XX
XX 5 PKW--PEPY-----RGLASDPGPEGEYANDQERRMTLAPPGYR 41
XX 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
XX 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
XX
XX 403 PTWDSKEPVCIACGAGVIRNATTCRIGVSPGFGNYSNNLTCHMLLEAPGQR 454
XX
XX RESULT 20
XX ID AAB70538 standard; Protein: 994 AA.
XX
XX AAB70538;
XX
XX 09-MAY-2001 (first entry)
XX
XX Human PRO8 protein sequence SEQ ID NO:16.
XX
XX Human; PRO; cytosolic; immunomodulatory; reproduction;
XX

```

KW gene therapy; cell proliferation; differentiation disorder; cancer;
 KW Immune associated disorder; gestational disease; pre-clampsia.
 XX
 OS Homo sapiens.
 XX
 PN WO200110902-A2.
 XX
 PD 15-FEB-2001.
 XX
 PE 11-AUG-2000; 2000WO-US21857.
 XX
 PR 11-AUG-1999; 99US-0148433.
 PR 10-AUG-2000; 2000US-0148433.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Fernandes E;
 DR WPI: 2001-147509/15.
 DR N-PSDB; AAF74439.
 XX
 XX Nucleic acids encoding secreted polypeptides, designated PROX
 PT polypeptides, useful for treating a syndrome associated with a
 PT PROX-associated disorder, e.g. cancer -
 XX
 OS Claim 1; Page 28-31; 16pp; English.

The present invention describes isolated nucleic acids encoding secreted polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide wherein X is an integer from 1 to 17). PROX polypeptides have cytostatic, immunomodulatory and reproduction activities, and can be used in gene therapy, and as PROX antagonists and PROX agonists. PROX polypeptides, nucleic acids and antibodies are useful in the manufacture of a medicament for treating a syndrome associated with a PROX-associated disorder, e.g. a cell proliferation and/or differentiation disorder (e.g. cancer or immune associated disorders) and a gestational disease (e.g. pre-eclampsia). They are also used for screening for a modulator of activity or of latency or predisposition to a PROX-associated disorder. AAT74432 to AAT74448 encode the specifically claimed human PROX polypeptides PRO1 to PRO17 given in AAT70531 to AAT70547.

```
SQ      Sequence    994 AA;                Score 91.5; DB 22; Length 994:
Query Match          38.4%;
Best Local Similarity 40.4%;       Pred. No. 0.0066;
Matches   21; Conservative     3; Mismatches  13; Indels  15; Gaps    2.

OY      5 PKW--PEPV-----EGRLASDFGFGXYANDDERRWTLTAPPGYR 41
           |||||
DB      403 PLWDSEKFEPCIIACGGVIRNATTGTGIVSGFPGNGSNLNTCHMLLEAPEGOR 454
```

Search completed: January 11, 2003, 10:48:09
Job time : 119 secs

RESULT	20
ID	AAB70538 standard; Protein: 994 AA.
XX	AAB70538;
XX	
AC	AAB70538;
XX	
DT	09-MAY-2001 (first entry)
XX	
DE	Human PR08 protein sequence SEQ ID NO:16
XX	
KW	Human; PRO; PROX; cytosolic; immunomodulatory

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 11, 2003, 11:06:49 : Search time 3185 Seconds

(without alignments)
374.636 Million cell updates/sec

Title: US-09-874-198-1
Perfect score: 238
Sequence: 1 TPLGKMPDPVFGRLASPGF.....GEYANDQERRWTLTAPPCVR 41

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Command line parameters:
-MODE=frame_plus.p2n.model -DEV=rlp
-Q/cgn2_l/uspt0.spool/us09874198/runat_10012003_092043_2562/app-query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=60
-DOCAIGN=200 -THR.SCORE=del -THR.MAX=100 -THR.MIN=0 -ALIGN=20 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us09874198.accn1.1.3637 -runat_10012003_092043_2562 -NCPU=6 -TCPU=3
-NO_XLPRY -NO_MMAP -LARGEBUFFER -NEG.SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGABEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_hlg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	238	100.0	683	9	HOSA18282	Y18282 Homo sapien
2	238	100.0	725	9	AB008047	AB008047 Homo sapien
3	238	100.0	729	9	HOSA18283	Y18283 Homo sapien
4	238	100.0	736	9	HOSA18281	Y18281 Homo sapien
5	238	100.0	1184	9	HOSA18284	Y18284 Homo sapien
6	238	100.0	2450	9	HSMASP2	X98400 H.sapiens m
7	238	100.0	2455	9	HSMASPPR	Y09926 Homo sapien
8	238	100.0	2819	9	AB033742	AB033742 Homo sapien
9	238	100.0	4464	9	HOSA18286	Y18286 Homo sapien
10	238	100.0	4465	9	HOSA18287	Y18287 Homo sapien
11	238	100.0	4900	9	AF321558S1	AF321558 Homo sapien
12	238	100.0	112769	9	HSJ635E18	AL109811 Human DNA
13	183	76.9	733	10	MMU250369	AJ7250369 Mus muscu
14	183	76.9	817	10	MMU19160	Y19160 Mus muscu
15	183	76.9	1145	10	MMU19163	Y19163 Mus muscu
16	183	76.9	2050	10	AB009459	AB009459 Mus muscu
17	183	76.9	3070	10	BC013893	BC013893 Mus muscu
18	183	76.9	205573	10	AL591032	AL591032 Mouse DNA
19	183	76.9	224312	2	AL606969	AL606969 Mus muscu
20	180	75.6	2037	10	RNO277747	AJ277747 Rattus no
21	180	75.6	172473	2	AC115553	AC115553 Rattus no
22	180	75.6	186184	2	AC094176	AC094176 Rattus no
23	170	71.4	1800	10	MMU19161	Y19161 Rattus norv
24	164	68.9	658	10	RNO18573	Y18573 Rattus Norv
25	164	68.9	680	10	RNO18570	Y18570 Rattus Norv
26	164	68.9	715	10	RNO18568	Y18568 Rattus Norv
27	100	42.0	2282	5	AB009072	AB009072 Xenopus l
28	93	39.1	449	9	AB01082502	AB010813 Homo sapi
29	93	39.1	636	10	AF004661	AF004661 Rattus no
30	93	39.1	1013	9	AB00760502	AB007603 Homo sapi
31	93	39.1	2108	10	RNO277423	AJ277423 Rattus no
32	93	39.1	2787	9	HUMMASP	D28593 Human mRNA
33	93	39.1	3736	10	AB049755	AB049755 Mus muscu
34	93	39.1	3863	6	AX254423	AX254423 Sequence
35	93	39.1	3895	6	AX156466	AX156466 Sequence
36	93	39.1	3895	9	AF284421	AF284421 Homo sapi
37	93	39.1	4489	9	D17525	D17525 Human mRNA
38	93	39.1	5135	10	MUSCRARF	D16492 Mouse mRNA
39	93	39.1	62047	2	AC117672	AC117672 Mus muscu
40	93	39.1	154160	2	AC046154	AC046154 Homo sapi
41	93	39.1	173534	9	AC007920	AC007920 Homo sapi
42	93	39.1	207841	2	AC072019	AC072019 Homo sapi
43	91.5	38.4	1988	6	AX084227	AX084227 Sequence
44	91.5	38.4	2095	9	AF502130	AF502130 Homo sapi
45	91.5	38.4	2127	6	AX084223	AX084223 Sequence
46	91.5	38.4	2127	6	AX084225	AX084225 Sequence
47	91.5	38.4	2143	6	AX084229	AX084229 Sequence
48	91.5	38.4	2559	6	AX327363	AX327363 Sequence
49	91.5	38.4	3470	9	AK096473	AK096473 Homo sapi
50	91.5	38.4	3828	9	AK091522	AK091522 Homo sapi

51 91.5 38.4 3863 6 AX084219 Sequence
 52 91.5 38.4 3879 6 AX084221 Sequence
 53 91.5 38.4 4034 6 AX402474 Sequence
 54 91.5 38.4 4047 9 HSM805489
 55 91.5 38.4 4198 6 AX327362 Sequence
 56 91.5 38.4 4227 6 AY038048 Homo sapi
 57 91.5 38.4 4306 9 AF502129 Homo sapi
 58 91 38.2 174701 2 AC024267 Homo sapi
 59 36.6 220638 10 AL645723 Mouse DNA
 60 85 35.7 33556 3 U00036 Caenorhabditis

ALIGNMENTS

RESULT 1
 HOSA18282 683 bp mRNA linear PRI 12-JUL-1999
 LOCUS Homo sapiens mRNA for mannose binding lectin-associated serine
 DEFINITION protease-2, alternatively spliced transcript (clone ph1-7).

Y18282
 Y18282.1 GI:5459316
 lectin-associated protein; mannose binding lectin-associated
 protein; mannose binding protein; serine protease.
 Homo sapiens.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,
 Jøensenius,O.C. and Schwaebie,W.J.
 TITLE Two constituents of the initiation complex of the mannan-binding
 JOURNAL lectin activation pathway of complement are encoded by a single
 structural gene
 J. Immunol. 162 (6), 3481-3490 (1999)

MEDLINE
 PUBMED 10092804
 2 (bases 1 to 683)
 Schwaebie,W.J.
 DIRECT SUBMISSION
 Submitted (19-OCT-1998) W.J. Schwaebie, Department of Microbiology
 and Immunology, University of Leicester, University Road, PO Box
 138, Leicester LE1 9HN, UK
 Related sequence Y09926.

COMMENT
 FEATURES
 source
 1. 683
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1p36.2-3"
 /clone="ph1-7"
 /tissue_type="liver"
 1. 683
 /gene="MASP-2"
 <1..528
 /gene="MASP-2"
 /note="alternative"
 /codon_start=1
 /product="mannose binding lectin-associated serine
 protease-2 related protein, MAP19 (19kDa)"
 /protein_id="CAB50729.1"
 /db_xref="GI:5459317"
 /translation="CGSVATPLGPKWPEVFGRLASPGFGEVANDQERRWTLTAAPG
 YRLRYFTHTDELSHLCEYDVKLSSGAKVLATCGESTDTERAPGKDFYSLGSS
 LITFRSDYSNKKPTGFEAFYAADIDECQVAPGFAFPCDHCHNHLCGFCSCNAG
 YLHRKRTCSQSL"
 <1..15
 /gene="MASP-2"
 16..525
 /gene="MASP-2"
 /product="mannose binding lectin-associated serine
 protease-2 related protein, MAP19 (19kDa)"
 /note="alternative"
 529..683

poly_a_signal
 BASE COUNT 136 a 246 c 187 g 114 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,45e-16 Length: 683
 Score: 238.00 Matches: 41
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HOSA18282 (1-683)

QY 1 ThThProLeuGlyProIySTrPrOgiuProValPheGlyYArgLeuAlaSerProgiuPhe 20
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 16 ACCCCCTTAGGCCCGAAGTGCCCTGAACCTGTGTGGGGCGCTGCATCCCGGCTTT 75

QY 21 ProGLyGLuTrAlaAsnAspGInGluArgTrPThrLeuThraLapProproGLyTYr 40
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 76 CCAAGGAGATGATCCATGACACGAGAGCGGCTGACCTGACTACACCCCGGCTTAC 135

QY 41 Arg 41
 |||
 Db 136 CGC 138

RESULT 2
 AB008047 725 bp mRNA linear PRI 08-JUN-1999
 LOCUS Homo sapiens sMAP mRNA for small MBL-associated protein, complete
 DEFINITION cds.

ACCESSION
 VERSION AB008047.1 GI:5002493
 KEYWORDS small MBL-associated protein; sMAP.
 SOURCE Homo sapiens cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Takahashi,M., Endo,Y., Fujita,T. and Matsushita,M.
 TITLE A truncated form of mannose-binding lectin-associated serine
 JOURNAL protease (MASP)-2 expressed by alternative polyadenylation is a
 component of the lectin complement pathway
 Int. Immunol. 11 (5), 859-863 (1999)

MEDLINE
 PUBMED 99262288
 2 (bases 1 to 725)
 Takahashi,M., Matsushita,M. and Fujita,T.
 DIRECT SUBMISSION
 Submitted (10-OCT-1997) Minoru Takahashi, Fukushima Medical
 College, Department of Biochemistry, 1 Hikarigaoka, Fukushima,
 Fukushima 960-12, Japan (E-mail:minoru@taecc.fmu.ac.jp,
 Tel:81-245-48-2111, Fax:81-245-48-2111)
 Sequence update (13-Dec-1997).

COMMENT
 FEATURES
 source
 1. 725
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..725
 /gene="sMAP"
 27..584
 /gene="sMAP"
 /codon_start=1
 /product="small MBL-associated protein"
 /protein_id="BAA78616.1"
 /db_xref="GI:5002494"
 /translation="MRLLTLGLCGSVATPLGPKWPEVFGRLASPGFGEVANDQERRWTLTAAPG
 KRWTLAPRGVRLRLYFTHTDELSHLCEYDVKLSSGAKVLATCGESTDTERAPG
 KRTYSLGSLDITFRSDYSNKKPTGFEAFYAADIDECQVAPGFAFPCDHCHNHLCGFCSCNAG
 GGFYSCSRAGYVLHRKRTCSQSL"
 27..71
 /gene="sMAP"

sig_peptide

mat_peptide 72..581
 /gene="sMAP"
 /product="unnamed"
 polyA_signal 706..711
 /gene="sMAP"

BASE COUNT 126 a 268 c 208 g 123 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 238.00	725	41	0	0	0	0
Percent Similarity: 100.00%						
Best Local Similarity: 100.00%						
Query Match: 100.00%						

US-09-874-198-1 (1-41) x AB008047 (1-725)

1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
 |||
 72 ACCCCCTGGGGCGGAAGTGGCTGAACTGTGTGGGGCGCTGGCATCCCCGCTTT 131

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProGlyTyr 40
 |||
 Db 132 CCAGGGAGATGTCATGACACAGAGCGCGCTGACCTGACTGCACCCCGGCTAC 191

QY 41 Arg 41
 |||
 Db 192 CGC 194

RESULT 3
 HOSA18283 729 bp mRNA linear PRI 12-JUL-1999

LOCUS

DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine protease-2, alternatively spliced transcript (clone phl-6).

ACCESSION Y18283

VERSION Y18283.1 GI:5459318

KEYWORDS lectin-associated protein; mannose binding lectin-associated protein; mannose binding protein; serine protease.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 729)
 Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T., Jensenius,J.C. and Schwaebler,W.J.
 Two constituents of the Initiation complex of the mannan-binding lectin activation pathway of complement are encoded by a single structural gene

JOURNAL J. Immunol. 162 (6), 3481-3490 (1999)

MEDLINE 99192764

PUBMED 10092804

REFERENCE 2 (bases 1 to 729)
 Schwaebler,W.J.
 Direct Submission
 Submitted (19-OCT-1998) W.J. Schwaebler, Department of Microbiology and Immunology, University of Leicester, University Road, PO Box 138, Leicester LE1 9HN, UK

COMMENT Related sequence Y09926.

FEATURES

source

1..729
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1p36.2-3"
 /clone="phl-6"
 /tissue_type="liver"
 1..720
 /gene="MASP-2"
 1..15
 /gene="MASP-2"
 16..573
 /gene="MASP-2"
 /gene="MASP-2"

/note="alternative"
 /codon_start=1
 /product="mannose binding lectin-associated serine protease-2 related protein, Map19 (19kDa)"
 /protein_id="CA50730.1"
 /db_xref="GI:5459319"
 /translation="MRLLTLGLLCSVATPLGPKPEVFGRLASPGFGEVANDOE RMTLTPAPGYRLRYFTHTDLESLICEDYDFVKLSGAKVLATLGGSTPTERAG KDFYSLGSSLDITFRSDYSNENRPTFGFAFAAEDIDECVAPRGAPICDHHCNHL GGFYCSGRAGYVLRNKRKTSQSL"
 16..60
 /gene="MASP-2"
 61..570
 /gene="MASP-2"
 /product="mannose binding lectin-associated serine protease-2 related protein, Map19 (19kDa)"
 574..720
 /gene="MASP-2"
 693..698
 /gene="MASP-2"

BASE COUNT 142 a 264 c 201 g 122 t

ORIGIN

polyA_signal

3'UTR

sig_peptide

mat_peptide

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 238.00	729	41	0	0	0	0
Percent Similarity: 100.00%						
Best Local Similarity: 100.00%						
Query Match: 100.00%						

US-09-874-198-1 (1-41) x HOSA18283 (1-729)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
 |||
 Db 61 ACCCCCTAGCGCCGGAAGTGGCTGAACTGTGTGGGGCGCTGGCATCCCCGCTTT 120

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProGlyTyr 40
 |||
 Db 121 CCAGGGAGATGTCATGACACAGAGCGCGCTGACCTGACTGCACCCCGGCTAC 180

QY 41 Arg 41
 |||
 Db 181 CGC 183

RESULT 4
 HOSA18281 736 bp mRNA linear PRI 12-JUL-1999

LOCUS

DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine protease-2, alternatively spliced transcript (clone phl-5).

ACCESSION Y18281

VERSION Y18281.1 GI:5459314

KEYWORDS lectin-associated protein; mannose binding lectin-associated protein; mannose binding protein; serine protease.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 736)
 Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T., Jensenius,J.C. and Schwaebler,W.J.
 Two constituents of the Initiation complex of the mannan-binding lectin activation pathway of complement are encoded by a single structural gene

JOURNAL J. Immunol. 162 (6), 3481-3490 (1999)

MEDLINE 99192764

PUBMED 10092804

REFERENCE 2 (bases 1 to 736)
 Schwaebler,W.J.
 Direct Submission
 Submitted (19-OCT-1998) W.J. Schwaebler, Department of Microbiology and Immunology, University of Leicester, University Road, PO Box 138, Leicester LE1 9HN, UK

REMARK Revised by author (17-JUL-1999)
COMMENT Related sequence Y09926.
FEATURES
source
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-3"
/clone="phl-5"
/tissue_type="liver"
1..736
/gene="MASP-2"
1..21
/gene="MASP-2"
22..579
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="mannose binding lectin-associated serine
protease-2 related protein, MAP19 (19kDa)"
/protein_id="CAB50728.1"
/db_xref="GI:5459315"
/translation="MRLLFLGLGGSVATPLGPKMPPEVFGRLASPGFGYANDOE
KRMFLAPPGYRLRYFTFHFDELSHCYDVKLSGAKVLATLGGDSTDERAPG
KDFYSLGSSLDITFRSDYSNEKPFGEAFYAAEDIDECQVAPGEAPTCDDHCHNHL
GGFYCSGRAGYVLRNKRKTCSEOSL"
581..736
/gene="MASP-2"
700..705
/gene="MASP-2"
BASE COUNT 144 a 266 c 203 g 123 t
ORIGIN
Alignment Scores:
Pred. No.: 6.93e-16 Length: 736
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-874-198-1 (1-41) x HOSA18281 (1-736)
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 67 ACCCCCTAGGCCGGAAGTGGCCTGAACCTGTGTGGGGCGCTGGCATCCCGGGCTTT 126
|||||
Ov 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrThrLeuThrAlaProProGlyTyr 40
|||||
127 CCAGGGGAGTATGCCAATGACGAGGCGGCGCTGACCTGACCTGACCCCGGCTAC 186
|||||
41 Arg 41
|||
Db 187 CGC 189
RESULT 5
LOCUS HOSA18284 1184 bp mRNA linear PRI 12-JUL-1999
DEFINITION Homo sapiens mRNA for mannose binding lectin-associated serine
protease-2, incompletely spliced primary transcript, clone phl-8.
ACCESSION Y18284
VERSION Y18284.1 GI:5459320
KEYWORDS lectin-associated protein; mannose binding lectin-associated
protein; mannose binding protein; serine protease.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1184)
Storer,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,
Jensenius,J.C. and Schwaebble,W.J.
.TITLE
lectin activation pathway of complement are encoded by a single

JOURNAL structural gene
J. Immunol. 162 (6), 3481-3490 (1999)
MEDLINE 99192764
PUBMED 10092804
REFERENCE 2 (bases 1 to 1184)
AUTHORS Schwaebble,W.J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology
and, Immunology, University of Leicester, University Road, PO Box
138, Leicester LE1 9HN, UK
COMMENT Related sequence Y09926
FEATURES
source
1..1184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-3"
/clone="phl-8"
/tissue_type="liver"
1..1184
/gene="MASP-2"
1..21
/gene="MASP-2"
22..564,1009..1023
/join(22..564,1009..1023)
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="mannose binding lectin-associated serine
protease-2 related protein, MAP19 (19kDa)"
/protein_id="CAB50731.1"
/db_xref="GI:5459321"
/translation="MRLLFLGLGGSVATPLGPKMPPEVFGRLASPGFGYANDOE
KRMFLAPPGYRLRYFTFHFDELSHCYDVKLSGAKVLATLGGDSTDERAPG
KDFYSLGSSLDITFRSDYSNEKPFGEAFYAAEDIDECQVAPGEAPTCDDHCHNHL
GGFYCSGRAGYVLRNKRKTCSEOSL"
22..66
/gene="MASP-2"
1024..1149
/gene="MASP-2"
1144..1149
/gene="MASP-2"
BASE COUNT 201 a 430 c 325 g 228 t
ORIGIN
sig_peptide
mat_peptide
3'UTR
polyA_signal
BASE COUNT 201 a 430 c 325 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 1.09e-15 Length: 1184
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-874-198-1 (1-41) x HOSA18284 (1-1184)
QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 67 ACCCCCTAGGCCGGAAGTGGCCTGAACCTGTGTGGGGCGCTGGCATCCCGGGCTTT 126
|||||
Ov 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrThrLeuThrAlaProProGlyTyr 40
|||||
127 CCAGGGGAGTATGCCAATGACGAGGCGGCGCTGACCTGACCTGACCCCGGCTAC 186
|||||
41 Arg 41
|||
Db 187 CGC 189
RESULT 6
LOCUS HSMASP2 2450 bp mRNA linear PRI 07-JUL-1998
HSMASP2

DEFINITION H.sapiens mRNA for mannan-binding lectin-associated serine
protease-2.
ACCESSION X98400
VERSION X98400.1 GI:3297878
KEYWORDS lectin-associated protein; mannan-binding protein; MASP; serine
protease.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2450)
Thiel,S., Vorup-Jensen,T., Stover,C.M., Schwaebler,W., Laursen,S.B.,
Poulsen,K., Willis,A.C., Eggleston,P., Hansen,S., Holmskov,U.,
Reid,K.B.M. and Jensenius,J.C.
Identification and characterization of a novel protein of the human
complement system, mannan-binding lectin-associated serine
protease-2 (MASP-2)
JOURNAL Unpublished
SOURCE 2 (bases 1 to 2450)
Thiel,S.
Direct Submission
Submitted (08-JUN-1996) S. Thiel, University of Aarhus, Dept. of
Medical Microbiology, Immunol., Bartholin Building, Wilhelm Meyers
Alle, 8000 Aarhus C, Denmark
LOCATION/Qualifiers
1. .2450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="4"
37. .2097
/note="MASP-2"
/codon_start=1
/product="mannan-binding lectin-associated serine
protease-2"
/protein_id="CA67050.1"
/db_xref="GI:3297878"
/translation="MRLTLGLGCGVATPLGPKMPPEVFGRLASPGFGEYANDOE
RRMTITAPGRLRLFTTHFDLESLHCEYDFVKLSAKYLATLTCGSESTDEAPG
KDTFYSLSGSDITFRSDYSNEKPTGEATYAADIDECOVAPEATCHHCHNHL
GGFYCSRCAGVYLHNRKRTCSALSGOYFTRSGELSSPEYPRPKLSSTCTYSISLE
EGFSVILDFVESFDETHPEFLCEYDFLKTIDTDEHGFPGKTLPHRIETKSNVTI
TFVDESQDHTGKIHVYSTAHACYPMAPPNGHVSPYQAKYILKDSFISICEYEL
LOGHLPLKSFPAVOKDGSMDRPMACSTVDCGPPDLPDSGRVEYITGPGVITYKAVI
OYSCERTFYTKVNDGKYVCADGFWTSKGEKSLPVCPEVCGLSARTTGGRTYGGOK
AKPGDFPMQVLLIGGTTAGALLDYNNVLTAAHAYEOKHDSALDITMGTLKRLSPH
YTQAMSEAVFTHEGYTHDAGFNDNALIKLNKKVYINNTITCLPRKEASFMTD
IGTASGWLGTORGFLARNLMYVDIPIVDHOKCTAAVEKPPYPRGSVTANMLCAGLESG
GKDCSRGDSGALVFLDSETEHMFVGGIVSGMNCBAGOGVYTVKVINIPIENT
ISDF"

BASE COUNT 609 a 644 c 617 g 580 t
ORIGIN

Alignment Scores:
Pred. No.: 2,19e-15 Length: 2450
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservation: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HSMASP2 (1-2450)

QY 1 ThrProLeuGlyProLyTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 82 ACCCCCTTAGGCGCCAGAGGCGCTGACCTGCTTCGGCGGCTGGACATCCCCGGCTTT 141
QY 21 ProGlyGluTyTAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyT 40
DB 142 CCAGGGGAGTATGCCAATGACAGAGCGCGCTGACCTGACCTGACACCCCCGGCTAC 201
QY 41 Arg 41
|||

DB 202 CGC 204
RESULT 7
HSMASP2PR 2455 bp mRNA linear PRI 12-DEC-1998
LOCUS Homo sapiens mRNA for MASP-2 protein.
DEFINITION Y09926
ACCESSION Y09926.1 GI:4007626
VERSION MASP-2 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2455)
Jensen,T.V.
Direct Submission
Submitted (06-DEC-1996) T.V. Jensen, University of Aarhus,
Microbiology & Immunology, Bartholin Building, 8000 Aarhus C,
DENMARK
REMARK Revised by [3]
REFERENCE 2 (bases 1 to 2455)
Thiel,S., Jensen,T.V., Stover,C.M., Schwaebler,W.J., Laursen,S.B.,
Poulsen,K., Willis,A.C., Eggleston,P., Hansen,S., Holmskov,U.,
Reid,K.B.M. and Jensenius,J.C.
A second serine protease associated with mannan-binding lectin that
activates complement
Nature 386 (6624), 506-510 (1997)
JOURNAL 9727412
MEDLINE 9087411
PUBMED 3 (bases 1 to 2455)
REFERENCE Schwaebler,W.J.
AUTHORS Direct Submission
TITLE Submitted (08-DEC-1998) W.J. Schwaebler, Department of Immunology
and Microbiology, University of Leicester, University Road,
Leicester LE1 9HN, UK
COMMENT On Dec 13, 1998 this sequence version replaced gi:1929053.
FEATURES
source
1. .2455
Location/Qualifiers
1. .2455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.2-3"
/tissue_type="liver"
1. .2455
/gene="MASP-2"
1. .16
/gene="MASP-2"
17. .2077
/gene="MASP-2"
/db_xref="SPTREMBL:000187"
/translation="MRLTLGLGCGVATPLGPKMPPEVFGRLASPGFGEYANDOE
RRMTITAPGRLRLFTTHFDLESLHCEYDFVKLSAKYLATLTCGSESTDEAPG
KDTFYSLSGSDITFRSDYSNEKPTGEATYAADIDECOVAPEATCHHCHNHL
GGFYCSRCAGVYLHNRKRTCSALSGOYFTRSGELSSPEYPRPKLSSTCTYSISLE
EGFSVILDFVESFDETHPEFLCEYDFLKTIDTDEHGFPGKTLPHRIETKSNVTI
TFVDESQDHTGKIHVYSTAHACYPMAPPNGHVSPYQAKYILKDSFISICEYEL
LOGHLPLKSFPAVOKDGSMDRPMACSTVDCGPPDLPDSGRVEYITGPGVITYKAVI
OYSCERTFYTKVNDGKYVCADGFWTSKGEKSLPVCPEVCGLSARTTGGRTYGGOK
AKPGDFPMQVLLIGGTTAGALLDYNNVLTAAHAYEOKHDSALDITMGTLKRLSPH
YTQAMSEAVFTHEGYTHDAGFNDNALIKLNKKVYINNTITCLPRKEASFMTD
IGTASGWLGTORGFLARNLMYVDIPIVDHOKCTAAVEKPPYPRGSVTANMLCAGLESG
GKDCSRGDSGALVFLDSETEHMFVGGIVSGMNCBAGOGVYTVKVINIPIENT
ISDF"

gene
5' UTR,
CDS
sig_peptide
3' UTR

CDS	join(30..34,116..344,501..678,1693..>1824)	
	/gene="MASP2"	
	/codon_start=1	
	/product="MBL-associated serine protease(MASP)-2"	
	/protein_id="BA85659.1"	
	/db_xref="GI:6092073"	
	/translation="MRLLTLGLGCGSVATPLGPKRPVFGRLASBPGEYANDOE RHWLTPAPGRLRLYPTTHFDLESHLCEYDFKLSGAKVLAITLGSESTDRPAPG KDPFYSIGSSLDITFRSDYSENERPFTGFAPFAAEDIDECQVAPGEAPTCDDHCNHL GGFSCSRAGVLAHRNRKTS"	
intron	35..115	/gene="MASP2"
	116..344	/gene="MASP2"
exon	/number=2	345..500
	/gene="MASP2"	501..678
exon	/gene="MASP2"	/number=3
	679..1692	/gene="MASP2"
intron	1693..1824	/gene="MASP2"
exon	/number=4	1825..2278
	/gene="MASP2"	2279..2433
intron	/gene="MASP2"	/note="specific exon for sMAP"
	/number=5	2414..2419
polyA_signal	/gene="MASP2"	2434..>2819
BASE COUNT	510 a 924 c 830 g 538 t 17 others	
ORIGIN		
Alignment Scores:		
Pred. No.:	2.5e-15	Length: 2819
Score:	238.00	Matches: 41
Percent Similarity:	100.00%	Conservative: 0
Best local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	9	Gaps: 0
US-09-874-198-1 (1-41) x ABO33742 (1-2819)		
Qy	1	ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
Db	156	ACCCCTTGGGGCCCAATGGCTGACCTGTGTTCCGGGCGCTGGCATCCCGGCTTT 215
Qy	21	ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40
Db	216	CCAGGGAGTATGCCAATGACACGAGCGCGCTGAGACCTGACTGCACCCCGGCTTAC 275
Qy	41	Arg 41
Db	276	CGC 278
RESULT 9		
LOCUS	HOSAI8286	4464 bp DNA linear PRI 12-JUL-1999
DEFINITION	Homo sapiens partial MASP-2 gene, clone pgm-2A.	
ACCESSION	Y18286	
KEYWORDS	GI:5459322	
VERSION	Y18286.1	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Stover,C.M., Thiel,S., Thelen,M., Lynch,N.J., Vorup-Jensen,T.,	

Jensenius, J.C. and Schwaebble, W.J.
Two constituents of the initiation complex of the mannan-binding
lectin activation pathway of complement are encoded by a single
structural gene
J. Immunol. 162 (6), 3481-3490 (1999)

JOURNAL
MEDLINE
99192764
PUBMED
10092804
REFERENCE
2 (bases 1 to 4464)
AUTHORS
Schwaebble, W.J.
TITLE
Direct Submission
JOURNAL
Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology
and, Immunology, University of Leicester, University Road, PO Box
138, Leicester LE1 9HN, UK

FEATURES
source
1. .4464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-3"
/clone="pgm-2A"
/cell_type="lymphocyte"
1. .4464
/gene="MASP-2"
join(1. .559, 2414. .2610, 2926. .4464)
/gene="MASP-2"
/note="alternative"
join(1. .559, 1004. .1156)
/gene="MASP-2"
/note="alternative"
1. .359
/gene="MASP-2"
/note="a"
join(16. .559, 2414. .2610, 2926. .4245)
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="MASP-2 protein"
/protein_id="CAB50733.1"
/db_xref="GI:5459324"
/translation="MRLTLILGLGSSVATPLGPKPEPVFGRLASPGPGEYANDOE
RKMULTAPPGYRLRLFTFLDELHSLCEYDFVKSAGAKVLATLGGESTDERAPG
KTFYSLGSSLDITERSDYSNEKPTFGFAFYAAEDIDECQVAPGAPPCDHNHNL
GGEYSCRAGYVILHNRKRTCSALCSGVFTORSGLSPSPRYPKLSCTYSILE
EGFSYILDFEVSFDEYTHPEETLCPYDFLKIQDREHGFEGCKTLPHRIETKSNVTI
TFTYDSGDHGMKIHVYSTANACPYMAPPNCHVSPNOAKYILKSPSEFCTGVEL
LQCHLPIKSFPAVCAQKDSMDRMPAIVYCCGPPDDLCSPGREGYITTGCVITTKAVI
QSCERTFTTKVNDGKIVCEADGFWTSSKGSILPCEVCGLSARTTGGRTYRGOK
AKPGDPWQVGLGTTAAGALLILYDMWVLTAAHAYEOKHDASALDIRGTLRLSPH
YQWAMEAVFIHEGYTHDAGFDNDIALILNNVYNSITPILCRKEAESPMRTD
IGTASGMGLTORGLARNLAMYDIPVDHOKCTAAVEKPPYPRGSGTANMLGLES
GDSCHGDSGALVFLDSETRFWVGIIYSWMSMNGEACQYCVITKYINIMIENTI
ISDF"
join(16. .559, 1004. .1017)
/gene="MASP-2"
/note="alternative"
/codon_start=1
/product="Mab19 protein"
/protein_id="CAB50732.1"
/db_xref="GI:5459323"
/translation="MRLTLILGLGSSVATPLGPKPEPVFGRLASPGPGEYANDOE
RKMULTAPPGYRLRLFTFLDELHSLCEYDFVKSAGAKVLATLGGESTDERAPG
KTFYSLGSSLDITERSDYSNEKPTFGFAFYAAEDIDECQVAPGAPPCDHNHNL
GGEYSCRAGYVILHNRKRTCSGSL"
1004. .1156
/gene="MASP-2"
/note="b"
2414. .2610
/gene="MASP-2"
/note="c"
2926. .4464
/gene="MASP-2"
/note="d"
1338 a 1217 g 971 t

CDS
938 a 1338 c 1217 g 971 t

ORIGIN
Alignment Scores:
Pred. No.: 3,87e-15 Length: 4464
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HOSA18286 (1-4464)

Qy 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 61 ACCCCCTTAGCCCGGAGCTGCGTGAACCTGTCTGGCGCCCTGGCATCCCGGCTTT 120
|||||
Qy 21 ProGluGlyTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40
|||||
Db 121 CCAGGGAGATGCGCATGACAGACGCGCGCTGACCTGACTGCACCCCGCGCTAC 180
|||||
Qy 41 Arg 41
|||||
Db 181 CGC 183

RESULT 10
HOSA18287 4465 bp DNA linear PRI 12-JUL-1999
HOSA18287
LOCUS
DEFINITION Homo sapiens partial MASP-2 gene, clone pgm-2B.
ACCESSION Y18287
VERSION Y18287.1 GI:5459325
KEYWORDS
protein; mannose binding protein; serine protease.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 4465)
Stover, C.M., Thiel, S., Thelen, M., Lynch, N.J., Vorup-Jensen, T.,
Jensenius, J.C. and Schwaebble, W.J.
Two constituents of the initiation complex of the mannan-binding
lectin activation pathway of complement are encoded by a single
structural gene
J. Immunol. 162 (6), 3481-3490 (1999)

JOURNAL
MEDLINE
99192764
PUBMED
10092804
REFERENCE
2 (bases 1 to 4465)
AUTHORS
Schwaebble, W.J.
TITLE
Direct Submission
JOURNAL
Submitted (19-OCT-1998) W.J. Schwaebble, Department of Microbiology
and, Immunology, University of Leicester, University Road, PO Box
138, Leicester LE1 9HN, UK

FEATURES
source
1. .4465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-3"
/clone="pgm-2B"
/cell_type="lymphocyte"
1. .4465
/gene="MASP-2"
join(1. .559, 2415. .2611, 2927. .4465)
/gene="MASP-2"
join(1. .559, 1004. .1156)
/gene="MASP-2"
1. .559
/gene="MASP-2"
/note="a"
join(16. .559, 2415. .2611, 2927. .4246)
/gene="MASP-2"
/codon_start=1
/product="MASP-2"
/protein_id="CAB50735.1"

CDS
10092804

DB:	9	Gaps:	0
Query Match:	100.00%	Mismatches:	0
Best local Similarity:	100.00%	Indels:	0
Score:	238.00	Conservative:	0
Alignment Scores:	3.88e-15	Matches:	4465
Protein No.:			
BASE COUNT	937 a 1338 c 1219 g 971 t		
ORIGIN			
exon	/note="b"		
exon	/gene="MASP-2"		
exon	/note="c"		
exon	/note="a"		
exon	/note="d"		
exon	/note="e"		
exon	/note="f"		
exon	/note="g"		
exon	/note="h"		
exon	/note="i"		
exon	/note="j"		
exon	/note="k"		
exon	/note="l"		
exon	/note="m"		
exon	/note="n"		
exon	/note="o"		
exon	/note="p"		
exon	/note="q"		
exon	/note="r"		
exon	/note="s"		
exon	/note="t"		
exon	/note="u"		
exon	/note="v"		
exon	/note="w"		
exon	/note="x"		
exon	/note="y"		
exon	/note="z"		
exon	/note="AA"		
exon	/note="AB"		
exon	/note="AC"		
exon	/note="AD"		
exon	/note="AE"		
exon	/note="AF"		
exon	/note="AG"		
exon	/note="AH"		
exon	/note="AI"		
exon	/note="AJ"		
exon	/note="AK"		
exon	/note="AL"		
exon	/note="AM"		
exon	/note="AN"		
exon	/note="AO"		
exon	/note="AP"		
exon	/note="AQ"		
exon	/note="AR"		
exon	/note="AS"		
exon	/note="AT"		
exon	/note="AU"		
exon	/note="AV"		
exon	/note="AW"		
exon	/note="AX"		
exon	/note="AY"		
exon	/note="AZ"		
exon	/note="BA"		
exon	/note="BB"		
exon	/note="BC"		
exon	/note="BD"		
exon	/note="BE"		
exon	/note="BF"		
exon	/note="BG"		
exon	/note="BH"		
exon	/note="BI"		
exon	/note="BJ"		
exon	/note="BK"		
exon	/note="BL"		
exon	/note="BM"		
exon	/note="BN"		
exon	/note="BO"		
exon	/note="BP"		
exon	/note="BQ"		
exon	/note="BR"		
exon	/note="BS"		
exon	/note="BT"		
exon	/note="BU"		
exon	/note="BV"		
exon	/note="BW"		
exon	/note="BX"		
exon	/note="BY"		
exon	/note="BZ"		
exon	/note="CA"		
exon	/note="CB"		
exon	/note="CC"		
exon	/note="CD"		
exon	/note="CE"		
exon	/note="CF"		
exon	/note="CG"		
exon	/note="CH"		
exon	/note="CI"		
exon	/note="CJ"		
exon	/note="CK"		
exon	/note="CL"		
exon	/note="CM"		
exon	/note="CN"		
exon	/note="CO"		
exon	/note="CP"		
exon	/note="CQ"		
exon	/note="CR"		
exon	/note="CS"		
exon	/note="CT"		
exon	/note="CU"		
exon	/note="CV"		
exon	/note="CW"		
exon	/note="CX"		
exon	/note="CY"		
exon	/note="CZ"		
exon	/note="DA"		
exon	/note="DB"		
exon	/note="DC"		
exon	/note="DD"		
exon	/note="DE"		
exon	/note="DF"		
exon	/note="DG"		
exon	/note="DH"		
exon	/note="DI"		
exon	/note="DJ"		
exon	/note="DK"		
exon	/note="DL"		
exon	/note="DM"		
exon	/note="DN"		
exon	/note="DO"		
exon	/note="DP"		
exon	/note="DQ"		
exon	/note="DR"		
exon	/note="DS"		
exon	/note="DT"		
exon	/note="DU"		
exon	/note="DV"		
exon	/note="DW"		
exon	/note="DX"		
exon	/note="DY"		
exon	/note="DZ"		
exon	/note="EA"		
exon	/note="EB"		
exon	/note="EC"		
exon	/note="ED"		
exon	/note="EE"		
exon	/note="EF"		
exon	/note="EG"		
exon	/note="EH"		
exon	/note="EI"		
exon	/note="EJ"		
exon			

AUTHORS	Park, D., Kim, B., Baek, K. and Yoon, J.
TITLE	Structure of Human MASP-2 Gene
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 4900)
AUTHORS	Park, D., Kim, B., Baek, K. and Yoon, J.
TITLE	Direct Submission
JOURNAL	Submitted (16-NOV-2000) Genetic Engineering, Kyung Hee University, 1 Seochun-Ri, Kihung-UP, Yongin City, Kyungki DO 449-701, Korea
FEATURES	Location/Qualifiers
source	1. 4900
organism	"Homo sapiens"
db_xref	"taxon:9606"
join	(1..172,254..482,639..816,1830..1961,2407..2859)
gene	"MASP2"
product	"MBL-associated protein Map19"
note	"alternatively spliced"
1..172	
gene	"MASP2"
number	1
join	(168..172,254..482,639..816,1830..1961,2407..2420)
gene	"MASP2"
note	"alternatively spliced"
codon_start	1
product	"MBL-associated protein Map19"
protein_id	"MAGS0275.1"
db_xref	"GI:12276137"
translation	"MRLTLGLGGSVATPLGPKPEVFGRLASPGFGEYANDQERMTLAPGVRRLRYFTYHPDLDSLHCEDYFVKSAGKVLATLGGESTPTERAPGKDFEYLSGSSLDITFRDYSNEKPTFGFEAFVAAEDIDECYVARGAPRCSHDHNCNHLGGFPCSRAGYVLIHRNKRTCSQSL"
254..482	
gene	"MASP2"
number	2
639..816	
gene	"MASP2"
number	3
1830..1961	
gene	"MASP2"
number	4
2407..2859	
gene	"MASP2"
note	"alternatively spliced"
number	5
3818..4014	
gene	"MASP2"
note	"alternatively spliced"
number	5
4328..4474	
gene	"MASP2"
number	6
1604	
1374	
982	
t	
BASE COUNT	940 a 1604 c 1374 g 982 t
ORIGIN	
Alignment Scores:	
Pred. No.:	4.24e-15 Length: 4900
Score:	238.00 Matches: 41
Percent Similarity:	100.00% Conservative: 0
Best local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	9 Gaps: 0
US-09-874-198-1 (1-41) x AF321558S1 (1-4900)	
Oy	1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
Db	294 ACCCCCTTAGGCCCGGAGTAGGCTGAACCTGTGTTCGGGGCCCTGGCATCCCCGGCTTT 333
Oy	21 ProGlyIuTyraIaasnAspGlnGluIuArgArgrTrpThrLeuThrAlaProProGlyTyr 40
Db	354 CAGGGAGGTATGCCAATGACAGGAGCGGGCGTGGACCTGACTGCACCCCGGGCTAC 413
Oy	41 Arg 41

Db 414 CGC 416

RESULT 12
LOCUS HSJ635E18/c
DEFINITION Human DNA sequence from clone RP4-635E18 on chromosome
1p36.11-36.31, complete sequence.
ACCESSION AL109811
VERSION AL109811.40 GI:181523762
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 112769)

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT SUBMISSION
Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jan 15, 2002 this sequence version replaced gi:11967852.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormep/
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP4-635E18 is from the library RPT-4 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

RES
source
1.112769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.11-36.31"
/clone="RP4-635E18"
/clone_1ib="RPT-4"
2691.3057
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
3696.36769
/note="Single clone region. Assembly confirmed by
restriction digest data. Single read sequenced with dGTP"
37125.37470
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
38084
/note="Forced join. Gap sized to be approximately 339bp
from restriction digest data and spanning pUC clone."
56138.56144
/note="sequence from overlapping clone d0576K7

(AL359082). Assembly confirmed by restriction digest
data."

BASE COUNT 29406 a 25975 c 28105 g 29283 t
ORIGIN

Alignment Scores:
Pred. No.: 8.46e-14 Length: 112769
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x HSJ635E18 (1-112769)

Oy 1 ThrProLeuAllyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 43427 ACCCCCTTGCGCCCGAAGTGGCTGACCTGTCTGCGCGCCCTGCGATCCCGGCTTT 43368
|||||

Oy 21 ProGlyGlyTyrAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProGlyTyr 40
|||||
Db 43367 CCAGGGAGATATGCCAATGACCAAGAGCGCGCTGGACCTGACTGCACCCCGGCTTAC 43308
|||||

Oy 41 Arg 41
|||

Db 43307 CGC 43305

RESULT 13
LOCUS MMU250369 733 bp mRNA linear ROD 26-JAN-2000
DEFINITION Mus musculus mRNA for mannose binding lectin-associated serine
protease-2 related protein, (MASP-2/Map19 gene).
ACCESSION AJ250369
VERSION AJ250369.1 GI:6599308
KEYWORDS mannose binding lectin-associated serine protease-2; MASP-2/Map19
gene.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 733)
TITLE The rat and mouse homologues of MASP-2 and Map19, components of the
lectin activation pathway of complement
J. Immunol. 163 (12), 6848-6859 (1999)
JOURNAL 20054576
MEDLINE 10586086
PUBMED 2 (bases 1 to 733)
REFERENCE
AUTHORS Stover C.M.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1999) Stover C.M., Dept. of Experimental
Immunology, University of Marburg, Deutschhausstrasse 2, 35037
Marburg, GERMANY
COMMENT This mouse Map19 mRNA species differs from mRNA in Y19160 in its
5' end.

FEATURES
source
1.733
/organism="Mus musculus"
/variety="C57Bl6"
/db_xref="taxon:10090"
1.31
/gene="mouse Map19"
1.31
/gene="mouse Map19"
/function="unknown"
32..589
/gene="MASP-2/Map19"
32..589
/gene="MASP-2/Map19"
/function="unknown"
/codon_start=1
/product="mouse Map19"
/protein_id="CAB63701.1"

```
/db_xref="GI:6599309"
/translation="MRLLFLGLMSLVATLIGSKMPEPVFGRLVSPGPEKADHD
RSMVTLAPPGYRLRYFTHFDELSTRCEDYDKVLSGKVLATLGGESTPDQAPG
NRTFYSLSGSLKVTFTHSDSYSNKRPPTGFPAFYAEDVDECRVSLGDSVPCDHYCHNYL
GGYGCSCRAGYVLHONKHTECSQSL"
sig_peptide
32..76
/gene="MASP-2/Map19"
77..586
/gene="MASP-2/Map19"
mat_peptide
350..733
/product="mouse Map19"
3'UTR
590..733
/gene="mouse Map19"
590..733
/function="unknown"
polyA_signal
718..723
/gene="mouse Map19"
BASE COUNT 163 a 226 c 182 g 162 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612

```

gene	1. .1145				
5'UTR	1. .44				
CDS	45. .>1145				
	/gene="MASP-2"				
	/codon_start=1				
	/product="mannose binding lectin-associated serine protease-2"				
	/protein_id="CAB65250.1"				
	/db_xref="GI:6686733"				
	/translation="MSLPQQLIFLGLMSLVATLLGSKMPEPFGRLSPGHPERYADHDRSMWTLTAPGYRLRLYFTFHDELSRCYDPAEDVDCRVSLDPSVCDHYEADPAGNTPFYSLGSLKVTPEHSDVSNKPFGEAFPAEDVDCRVSLDPSVCDHYCHNLTGEGYCSAGAGYVLHONKHTGSLGSGVPTGKSYLSPEYPPYPKLISCTYSIRLEDFSVILDPVEFEDVETPEACCPDLSLKTQYDKGEHGFCKTLPRIETDSHKTTITRAIDSGNHTGMKTHYSTARCPDPTAPPNGSISPVQAIYVLKDRYFVCKTGFELLGGSVPLKSFYAVCQKDGSDMRPMPCSS"				
	45. .104				
sig_peptide					
COUNT	256 a	345 c	288 g	256 t	
ALIGNMENT					
Alignment Scores:					
Pred. NO.:	4.55e-10	Length:	1145		
Score:	183.00	Matches:	32		
Percent Similarity:	85.37%	Conservative:	3		
Best local Similarity:	78.05%	Mismatches:	6		
Query Match:	76.89%	Indels:	0		
DB:	10	Gaps:	0		
US-09-874-198-1 (1-41) x MMU19163 (1-1145)					
OY	1	ThrProLeuGlyProLyTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe	20		
Db	105	ACACTTCGCGGTTCACAAATGGCCCTAACCTGATTCGGGCGCTGGTCCCTGGCTTC	164		
OY	21	ProGlyGlyTrpAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr	40		
Db	165	CCAGAGAGATGCTGCACCATCAAGATGCATGCTGACACACACACCCCTGGCTAC	224		
OY	41	Arg	41		
Db	225	CGC	227		
RESULT 16					
AB009459	2090 bp	mRNA	linear	ROD 13-FEB-1999	
AB009459					
AB009459.1	GI:3928516				
manose-binding lectin associated serine protease-2.					
Mus musculus (strain: BALB/c) 8 weeks liver cDNA to mRNA.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
Endo. Y., Takahashi, M., Nakao, M., Saiga, H., Sekine, H., Matsushita, M., Nonaka, M. and Fujita, T.					
Two lineages of manose-binding lectin-associated serine protease (MASP) in vertebrates					
J. Immunol. 161 (9), 4924-4930 (1998)					
2 (bases 1 to 2090)					
Takahashi, M., Fujita, T., Endo, Y. and Matsushita, M.					
Submitted (04-DEC-1997) Minoru Takahashi, Fukushima Medical College, Department of Biochemistry, 1 Hikarigaoka, Fukushima, Japan (E-mail: minolta@cc.fmu.ac.jp, Tel: 81-24-548-2111, Fax: 81-24-548-6760)					
Location/Qualifiers					
1. .2090					

BASE COUNT	522 a	548 c	537 g	483 t
ORIGIN	SNP			
Alignment Scores:				
Pred. NO.:	8.08e-10			
Score:	183.00			
Percent Similarity:	85.37%			
Best Local Similarity:	78.05%			
Query Match:	76.89%			
DB:	10 Gaps:			
	0			
US-09-874-198-1 (1-41) x AB009459 (1-2090)				
Oy	1 ThrProLeaGlyProLysTrpProGluProValaPheGlyArgLeuAlaSerProGlyPhe 20			
Db	78 ACACCTCTGGGGTTCAAAATGGCTGTAACCTGATTCGGGCGCTGCTGCCCCCTGGCTTC 137			
Oy	21 ProGlyGluTrpAlaAspGlnuArGTrpThrLeuThAlaProProGlyTrp 40			
Db	138 CCAGGAGAGTATGCTGCACCATCAACATGATGCTGTGCACACAGCCCTGGCTAC 197			
Oy	41 Arg 41			
Db	198 CGC 200			
RESULT 17				
LOCUS	BC013893 3070 bp mRNA linear ROD 07-AUG-2002			
DEFINITION	Mus musculus, similar to mannan-binding lectin serine protease 2, clone MGC:13718 IMAGE:4211293, mRNA, complete cds.			
ACCESSION	BC013893			
VERSION	BC013893.1 GI:15530224			
KEYWORDS	MGC.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 3070)			
AUTHORS	Strausberg,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk			


```

RESULT 19
AL0606969      224312 bp      DNA      linear      HTG 13-AUG-2002
DEFINITION
Mus musculus chromosome 4 clone RP23-331P21, *** SEQUENCING IN
PROGRESS ***. 2 unordered pieces.
ACCESSION
AL0606969
VERSION
AL0606969.23      GI:22265372
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
house mouse.
ORGANISM
Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eumalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 224312)
Ramsay, H.
Direct Submission
Submitted (12-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humguy@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 15, 2002 this sequence version replaced gi:22204261.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humguy@sanger.ac.uk
----- Project Information
Center project name: BM331P21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 0% of reads
Sequencing vector: plasmid; 108752; 99% of reads
Chemistry: dye-terminator; 6% of reads
Chemistry: dye-terminator; Big Dye; 93% of reads
Consensus quality: 224154 bases at least Q40
Consensus quality: 224173 bases at least Q30
Consensus quality: 224186 bases at least Q20
Insert size: 224212; sum-of-contigs
Insert size: 221458; 2.4% error; agarose-fp
Quality coverage: 20.42x in Q20 bases; sum-of-contigs quality
coverage: 21.08x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 178012: contig of 178012 bp in length
* 178013 178112: gap of 100 bp
* 178113 224312: contig of 46200 bp in length.
Location/Qualifiers
1. 224312
/organism="Mus musculus"
/db_xref="taxon:100960"
/chromosome="4"
/clone="RP23-331P21"
/clone_1lb="RPCI-23"
1. 178012
/note="assembly_fragment:07594
fragment_chain:1"
178113. 224312
/note="assembly_fragment:01204
fragment_chain:1"
BASE COUNT      55283 a 55510 c 55759 g 57660 t 100 others
ORIGIN
Alignment Scores:
Pred. No.:      7.02e-08      Length:      224312
Score:          183.00      Matches:      32
Percent Similarity: 85.37%      Conservative: 3
Best Local Similarity: 78.05%      Mismatches: 6
Query Match:    76.89%      Indels:      0

```

DB: 2 Gaps: 0

US-09-874-198-1 (1-41) x ALG06969 (1-224312)

Oy 1 ThrProLeuGLyProIysTrpProGluProValPheGIlyArgLeuAlaSerProGlyPhe 20
||| |||||| | ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 22690 ACACCTTCGGGTGCAAAGTGGCCTCAACTGATTCGGGGCGCTGTGTCCTTGCCCTTC 22749

Oy 21 ProGlyIGluTyRAlaAsnAspSclInGuArGaTrPrhLeuThralaProProGlyITyr 40
||| ::::::::::::::: ||:::|| | ||||||| ||||||| ||||||| ||||||| |||||||
Db 22750 CCAGAAGATGATGCTGACCATCAAGATGATCTGTGACACTGACACTGCCCTGTCTAC 22809

Oy 41 Arg 41
|||
Db 22810 CGC 22812

RESULT 20
LOCUS RNO277747 2037 bp mRNA linear ROD 05-Oct-2000
DEFINITION Rattus norvegicus partial mRNA for mannose-binding protein associated serine protease-2 (masp-2 gene).
ACCESSION AJ277747
VERSION AJ277747.1 GI:7799288
KEYWORDS mannose-binding protein associated serine protease-2; MASP-2 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2037)
Wallis,R. and Dodd,R.B.
Intersection of mannose-binding protein with associated serine proteases: effects of naturally occurring mutations
J. Biol. Chem. 275 (40), 30962-30969 (2000)
MEDLINE 20469449
PUBMED 10913141
REFERENCE 2 (bases 1 to 2037)
Wallis,R.
Direct Submission
Submitted (06-MAY-2000) Wallis R., Department of Biochemistry,
Glycobiology Institute, University of Oxford, South Parks Road,
Oxford OX1 3OU, UNITED KINGDOM
FEATURES
source location/Qualifiers
1..2037
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_lib="liver"
1..2037
/gene="masp-2"
/gene="masp-2"
/function="complement activation"
/codon_start=1
/product="mannose-binding protein associated serine protease-2"
/protein_id="CAB90832.1"
/db_xref="gi:7799289"
translation="GLMSLVATLLGSNMPEVPGRLVSPFEPKYGNDHRSWTLTA
PGRRRLTFHFMLLELSTRCEDPVFKLTSGTKLATLCGESEIDTERAPNDPFYSI
GPSILKVFTHSDYISNEKPFTGEARVAABDECTSLDSSVPCHYCHNIUGYVCSC
RVGLITLQNKHTCSALCSGOVFTSRSEFLSSPERPOPYPKLSSCAVNI RLIEGFSITLD
DVESFDVEHMEPAOCYDSLKIQTDRREYPCFGKITLPRIETDSNKNVLTPTTDES
GNHGFWKITHYSTAOPCEDPTAPPNGHISTPVOATYVLADSVFCCKTELGQSVDPL
KSFLAVCOKDGSMDRP1PECSIIDCGPPDDL PNGAHVDITGPETTYKAVIAYSCEERT
PYTMSSNKKYICAEADCFWTSSKEKSLPYCKRPVCGHLSHTSGRTIGCPARKPDFFW
OVLLGETTAGALIHDDWVLTAAHAVYGTKEAASSLDIRNGILLRLSPHTOMPEPA
VTEHGTYTGAGFNDIALIKLRKVIYNRMIMPICLRKEAASLMKTDFGVYAAGWA
LTQGFGLARNLMFEVDPIVDHOKCATAYTKPOPGAKVTVMCLAGMCDAGCKDSCRDD
SGCALFLDNETQRMFVGIVSMGINSICGSEGYGVYTKVNTYIPMIENIIINE"
1..36
/gene="masp-2"
37..2034
/gene="masp-2"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 11, 2003, 12:10:24 : Search time 83 Seconds
(without alignments)
217.400 Million cell updates/sec

Title: US-09-874-198-1
Perfect score: 238
Sequence: 1 TPLGPKMPEVFGRLASPGF.....GEVANDQERMTLTAPGVR 41

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Sequences: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09874198/runat_10012003_092045_2707/app.query.fasta_1.199
-DB=Published_Applications_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=60 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100
-THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09874198.ecgn.1.1.36 @runat_10012003_092045_2707
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MAP -LARGEOVERT -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	2475	10 US-09-874-198-3	Sequence 3, Appl1
2	238	100.0	2475	10 US-09-874-238-3	Sequence 3, Appl1
3	93.5	39.3	810	10 US-09-764-853-143	Sequence 143, App
4	91.5	38.4	742	10 US-09-833-381-1924	Sequence 1924, Ap

ALIGNMENTS

RESULT 1
US-09-874-198-3
Sequence 3, Application US/09874198
Patent No. US20020082208A1
GENERAL INFORMATION:
APPLICANT: Jenselius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
TITLE OF INVENTION: USES FOR IT
FILE REFERENCE: 09011-002002
CURRENT APPLICATION NUMBER: US/09/874,198
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218

5	91.5	38.4	1988	9 US-10-004-551-21	Sequence 21, Appl
6	91.5	38.4	2127	9 US-10-004-551-17	Sequence 17, Appl
7	91.5	38.4	2127	9 US-10-004-551-19	Sequence 19, Appl
8	91.5	38.4	2143	9 US-10-004-551-23	Sequence 23, Appl
9	91.5	38.4	3863	9 US-10-004-551-13	Sequence 13, Appl
10	91.5	38.4	3879	9 US-10-004-551-15	Sequence 15, Appl
11	83.5	35.1	4344	10 US-09-917-800A-1712	Sequence 1712, Ap
12	81	34.0	1148	10 US-09-925-302-168	Sequence 168, Ap
13	81	34.0	1480	10 US-09-918-487-39	Sequence 39, Appl
14	81	34.0	1480	10 US-09-880-107-2196	Sequence 2196, Ap
15	81	34.0	2386	9 US-09-808-602-92	Sequence 92, Appl
16	81	34.0	2493	10 US-09-880-107-2256	Sequence 2256, Ap
17	81	34.0	2555	12 US-10-044-090-613	Sequence 613, App
18	80	33.6	354	10 US-09-960-352-3224	Sequence 3224, Ap
19	80	33.6	354	10 US-09-960-352-8661	Sequence 8661, Ap
20	80	33.6	425	10 US-09-960-352-10119	Sequence 10119, A
21	80	33.6	433	10 US-09-960-352-14174	Sequence 14174, A
22	79.5	33.4	416	10 US-09-960-352-1998	Sequence 1998, Ap
23	78.5	33.0	344	10 US-09-960-352-14817	Sequence 14817, A
24	78	32.8	390	10 US-09-960-352-5362	Sequence 5362, Ap
25	78	32.8	456	10 US-09-833-381-1925	Sequence 1925, Ap
26	78	32.8	559	10 US-09-833-381-1926	Sequence 1926, Ap
27	78	32.8	1542	10 US-09-822-849A-273	Sequence 273, App
28	78	32.8	1580	10 US-09-925-301-440	Sequence 440, App
29	78	32.8	1992	10 US-09-729-674-91	Sequence 91, Appl
30	76	31.9	18385	10 US-09-764-860-1018	Sequence 1018, Ap
31	74.5	31.3	691	10 US-09-925-300-416	Sequence 416, App
32	74.5	31.3	2647	10 US-09-880-107-2105	Sequence 2105, Ap
33	74.5	31.3	2659	12 US-10-044-090-472	Sequence 472, App
34	74.5	31.3	2725	10 US-09-925-301-182	Sequence 182, App
35	74	31.1	1547	10 US-09-917-800A-1668	Sequence 1668, Ap
36	70.5	29.6	414	10 US-09-960-352-4891	Sequence 4891, Ap
37	70	29.4	3955	10 US-09-855-722-4	Sequence 4, Appl1
38	70	29.4	4702	10 US-09-880-107-1545	Sequence 1545, Ap
39	69	29.0	2457	10 US-09-850-048A-1	Sequence 1, Appl1
40	69	29.0	3737	10 US-09-747-371-1	Sequence 1, Appl1
41	68	28.6	828	9 US-09-732-242-1	Sequence 1, Appl1
42	68	28.6	1366	10 US-09-800-729-15	Sequence 15, Appl
43	68	28.6	1440	12 US-10-067-422-6	Sequence 20, Appl
44	68	28.6	1440	10 US-10-098-841-207	Sequence 6, Appl1
45	68	28.6	1635	9 US-09-808-602-18	Sequence 207, App
46	68	28.6	1691	9 US-09-808-602-18	Sequence 16, Appl
47	68	28.6	3318	9 US-09-808-602-91	Sequence 18, Appl
48	68	28.6	3546	10 US-09-850-048A-3	Sequence 31, Appl
49	68	28.6	3919	10 US-09-942-366-4	Sequence 3, Appl1
50	68	28.6	5021	10 US-09-285-385C-1	Sequence 4, Appl1
51	67.5	28.4	2908	10 US-09-917-800A-1348	Sequence 1, Appl1
52	67.5	28.4	4661	10 US-09-285-385C-3	Sequence 1348, Ap
53	67	28.2	4771	10 US-09-942-366-2	Sequence 3, Appl1
54	66	27.7	684	10 US-09-833-381-1923	Sequence 2, Appl1
55	66	27.7	1127	10 US-09-764-853-132	Sequence 1923, Ap
56	65.5	27.5	231	10 US-09-964-824A-275	Sequence 132, App
57	65.5	27.5	1806	10 US-09-800-729-75	Sequence 275, App
58	65.5	27.5	1806	12 US-10-067-422-5	Sequence 75, Appl
59	65.5	27.5	2026	9 US-09-905-291A-103	Sequence 5, Appl1
60	65.5	27.5	2026	9 US-09-905-291A-103	Sequence 103, App

PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 2475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (37)...(2094)
US-09-874-198-3

Alignment Scores:

Pred. No.:	1,22e-20	Length:	2475
Score:	238.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Match:	100.00%	Indels:	0
	10	Gaps:	0

874-198-1 (1-41) x US-09-874-198-3 (1-2475)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||
DB 82 ACCCCCTTAGGCGGCGAGTGGCTGACCTGTGTGGCGGCTGACATCCCGGCTTT 141
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrPheThrAlaProProGlyTyr 40
|||
DB 142 CCAGGGAGTATGCCAATGACAGGCGCGCTGACCTGACCTGACCCCGGCTAC 201
QY 41 Arg 41
|||
DB 202 CGC 204

RESULT 2

US-09-874-238-3
Sequence 3, Application US/09874238 - with David Saunders
Patent No. US20020082209A1

GENERAL INFORMATION:

APPLICANT: Jensenius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
TITLE OF INVENTION: USES FOR IT
FILE REFERENCE: 09011-002003
CURRENT APPLICATION NUMBER: US/09/874,238
CURRENT FILING DATE: 2001-06-04
OR APPLICATION NUMBER: 09/054,218
OR FILING DATE: 1998-04-02
OR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 2475

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (37)...(2094)

US-09-874-238-3

Alignment Scores:

Pred. No.:	1,22e-20	Length:	2475
Score:	238.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-874-198-1 (1-41) x US-09-874-238-3 (1-2475)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||
DB 82 ACCCCCTTAGGCGGCGAGTGGCTGACCTGTGTGGCGGCTGACATCCCGGCTTT 141
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTyrPheThrAlaProProGlyTyr 40
|||
DB 142 CCAGGGAGTATGCCAATGACAGGCGCGCTGACCTGACCTGACCCCGGCTAC 201
QY 41 Arg 41
|||
DB 202 CGC 204

RESULT 3

US-09-764-853-143/C
Sequence 143, Application US/09764853
Patent No. US20020090672A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P3206
CURRENT APPLICATION NUMBER: US/09/764,853
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 939
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 143
LENGTH: 810
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-853-143

Alignment Scores:

Pred. No.:	0.00272	Length:	810
Score:	93.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	39.29%	Indels:	15
DB:	10	Gaps:	2

US-09-874-198-1 (1-41) x US-09-764-853-143 (1-810)

QY 5 ProLysTrp-----ProGluProVal----- 11
|||
DB 606 CCCWCTGGAGATTCAAAGAGGCCCGCTGTCATCGCTGTGGCGGAGTGTCCGCAT 547
QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
|||
DB 546 GCCACACCGCGCGCATCTCTCCAGGCTTCCGCGCACTACAGCAACACTCACC 487
QY 30 ArgArgTyrPheThrAlaProProGlyTyrArg 41
|||
DB 486 TGTCACTGGCTGCTTGAGGCTCTCTGAGGCGCAGCGG 451

RESULT 4

US-09-833-381-1924/C
Sequence 1924, Application US/09833381
Patent No. US20020132090A1

GENERAL INFORMATION:

APPLICANT: Robison, Keith E.
TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
FILE REFERENCE: 5800-119
CURRENT APPLICATION NUMBER: US/09/833,381
CURRENT FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 09/516,448
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 2050
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1924
LENGTH: 742
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1)...(742)
OTHER INFORMATION: n - A,T,C or G
US-09-833-381-1924

Alignment Scores:

Pred. No.: 0.00434 Length: 742
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
DB: 10 Gaps: 2

US-09-874-198-1 (1-41) x US-09-833-381-1924 (1-742)

QY 5 ProlystP-----ProgluProval----- 11

DB 606 CCCATCTGGGATTCGAAGAGCCGCTGTCATCGCTGCGGCGAGTGATCCGAAT 547

12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyrrAlaAsnAspGlnGlu 29

546 GCCACACCGCGCGCATCGTCTCCAGGCTCCCGGCGCAACTACAGCAACACCTCACC 487

QY 30 ArgArgTrrPThrLeuThrAlaProProGlyTyrrArg 41

DB 486 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 451

RESULT 5

US-10-004-551-21

Sequence 21, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

PRIOR FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 21

LENGTH: 1988

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)...(1752)

004-551-21

Alignment Scores:

Pred. No.: 0.0136 Length: 1988

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-21 (1-1988)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyrrAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCGCAACTACAGCAACACCTCACC 1503

QY 30 ArgArgTrrPThrLeuThrAlaProProGlyTyrrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

RESULT 6

US-10-004-551-17

Sequence 17, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

PRIOR FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 17

LENGTH: 2127

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)...(1752)

OTHER INFORMATION: n 2077 can be A, G, C, or T.

US-10-004-551-17

Alignment Scores:

Pred. No.: 0.0147 Length: 2127

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-17 (1-2127)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyrrAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCGCAACTACAGCAACACCTCACC 1503

QY 30 ArgArgTrrPThrLeuThrAlaProProGlyTyrrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

RESULT 7

US-10-004-551-19

Sequence 19, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

PRIOR FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 19

LENGTH: 2127

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)...(1752)

OTHER INFORMATION: n 2077 can be A, G, C, or T.

US-10-004-551-19

Alignment Scores:

Pred. No.: 0.0147 Length: 2127

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-17 (1-2127)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyrrAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCGCAACTACAGCAACACCTCACC 1503

QY 30 ArgArgTrrPThrLeuThrAlaProProGlyTyrrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

RESULT 8

US-10-004-551-19

Sequence 19, Application US/10004551

Publication No. US20030004310A1

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

PRIOR FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 19

LENGTH: 2127

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)...(1752)

OTHER INFORMATION: n 2077 can be A, G, C, or T.

US-10-004-551-19

Alignment Scores:

Pred. No.: 0.0147 Length: 2127

Score: 91.50 Matches: 21

Percent Similarity: 46.15% Conservative: 3

Best Local Similarity: 40.38% Mismatches: 13

Query Match: 38.45% Indels: 15

DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-17 (1-2127)

QY 5 ProlystP-----ProgluProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGAGCCGCTGTCATCGCTTGGCGGCGAGTGATCCGAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLeuTyrrAlaAsnAspGlnGlu 29

DB 1444 GGCACACCGCGCGCATCGTCTCCAGGCTCCCGGCGCAACTACAGCAACACCTCACC 1503

QY 30 ArgArgTrrPThrLeuThrAlaProProGlyTyrrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTGAGGCGCACGCG 1539

RESULT 9

Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 13
Best Local Similarity: 40.38% Mismatches: 3
Query Match: 38.45% Indels: 15
DB: 9 Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-19 (1-2127)

OY 5 ProlystP-----ProgiuProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGGAGCCCGTCTGCATGCTTGGCGGCGAGTGATCCGCAAT 1443

OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLysIleuAlaAsnAspGlnGlu 29

DB 1444 GGCACACACCGCGCGCATGCTCTCCAGGCTCCCGGCGCAACTACAGACACACCTCACC 1503

OY 30 ArgArgTrpThrLeuThrAlaProProGlyTyrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCGCACGCG 1539

US-10-004-551-23

Sequence 23, Application US/10004551

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

FILE REFERENCE: 15966-559

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 2143

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(1755)

OTHER INFORMATION: n 2093 can be A, G, C, or T.

US-10-004-551-23

Alignment Scores:

Pred. No.: 0.0149

Score: 91.50

Percent Similarity: 46.15%

Best Local Similarity: 40.38%

Query Match: 38.45%

DB: 9

Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-23 (1-2143)

OY 5 ProlystP-----ProgiuProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGGAGCCCGTCTGCATGCTTGGCGGCGAGTGATCCGCAAT 1443

OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLysIleuAlaAsnAspGlnGlu 29

DB 1444 GGCACACACCGCGCGCATGCTCTCCAGGCTCCCGGCGCAACTACAGACACACCTCACC 1503

OY 30 ArgArgTrpThrLeuThrAlaProProGlyTyrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCGCACGCG 1539

RESULT 9

US-10-004-551-13

Sequence 13, Application US/10004551

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

APPLICANT: FERNANDES, ELMA

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 13

LENGTH: 3863

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(3156)

OTHER INFORMATION: n 1755 can be A, G, C, or T

US-10-004-551-13

Alignment Scores:

Pred. No.: 0.0296

Score: 91.50

Percent Similarity: 46.15%

Best Local Similarity: 40.38%

Query Match: 38.45%

DB: 9

Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-13 (1-3863)

OY 5 ProlystP-----ProgiuProval----- 11

DB 1384 CCCTTCTGGGATTCGAAGGAGCCCGTATGCATGCTTGGCGGCGAGTGATCCGCAAT 1443

OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLysIleuAlaAsnAspGlnGlu 29

DB 1444 GGCACACACCGCGCGCATGCTCTCCAGGCTCCCGGCGCAACTACAGACACACCTCACC 1503

OY 30 ArgArgTrpThrLeuThrAlaProProGlyTyrArg 41

DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCGCACGCG 1539

RESULT 10

US-10-004-551-15

Sequence 15, Application US/10004551

GENERAL INFORMATION:

APPLICANT: SHIMKETS, RICHARD A

FILE REFERENCE: 15966-559

CURRENT APPLICATION NUMBER: US/10/004,551

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 09/635,949

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 110

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 15

LENGTH: 3879

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178)..(3159)

OTHER INFORMATION: n 1755 can be A, G, C, or T.

US-10-004-551-15

Alignment Scores:

Pred. No.: 0.0296

Score: 91.50

Percent Similarity: 46.15%

Best Local Similarity: 40.38%

Query Match: 38.45%

DB: 9

Gaps: 2

US-09-874-198-1 (1-41) x US-10-004-551-15 (1-3879)

QY 5 ProlystP-----ProgluProva]----- 11

Db 1384 CCCACTGGGATTCAAGAGCCGATGATGCTGCTGGCGGAGTGAATCCGCAAT 1443

QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAspGlnGlu 29

Db 1444 GGCACACCGCGCCGATCTCTCTCCAGGCTTCCGCGCACTACAGCAACACTCACC 1503

QY 30 ArgArgTrpThrLeuThraAlaProProGlyTyArg 41

Db 1504 TGTCACTGGCTGCTTGAGCTCTCCTGAGCGCCAGCGG 1539

RESULT 11

US-09-917-800A-1712

Sequence 1712, Application US/09917800A

PATENT No. US20020119462A1

GENERAL INFORMATION:

APPLICANT: Mendrick, Donna

APPLICANT: Porter, Mark

APPLICANT: Johnson, Kory

APPLICANT: Castle, Arthur

APPLICANT: Elashoff, Michael

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Molecular Toxicology Modeling

FILE REFERENCE: 44921-5038-US

CURRENT APPLICATION NUMBER: US/09/917.800A

PRIOR FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: US 60/222,040

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: US 60/222,880

PRIOR FILING DATE: 2000-11-02

PRIOR APPLICATION NUMBER: US 60/290,029

PRIOR FILING DATE: 2001-05-11

PRIOR APPLICATION NUMBER: US 60/290,645

PRIOR FILING DATE: 2001-05-15

PRIOR APPLICATION NUMBER: US 60/292,336

PRIOR FILING DATE: 2001-05-22

PRIOR APPLICATION NUMBER: US 60/295,798

PRIOR FILING DATE: 2001-06-06

PRIOR APPLICATION NUMBER: US 60/297,457

PRIOR FILING DATE: 2001-06-13

PRIOR APPLICATION NUMBER: US 60/298,884

PRIOR FILING DATE: 2001-06-19

PRIOR APPLICATION NUMBER: US 60/303,459

DR FILING DATE: 2001-07-09

DR OF SEQ ID NOS: 1740

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1712

LENGTH: 4344

TYPE: DNA

ORGANISM: Rattus norvegicus

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_022849

US-09-917-800A-1712

Alignment Scores:

Pred. No.: 0.331 Length: 4344

Score: 83.50 Matches: 16

Percent Similarity: 41.18% Conservative: 5

Best Local Similarity: 31.37% Mismatches: 17

Query Match: 35.08% Indels: 13

DB: 10 Gaps: 1

US-09-874-198-1 (1-41) x US-09-917-800A-1712 (1-4344)

QY 4 GlyProLysTrpProGluProAlaPhe----- 12

Db 2566 GGTCCAGTTTGGACAGTCTCTTTGTAACATACTTGTGAGGCTTCTCTGACTGGA 2625

QY 13 -----GlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAspGlnGluArg 30

Db 2626 CTCCTCGGGCAATTTTCTAGCCCATCTACCCCTGGAGCATCTATATATGCGAGATGT 2685

QY 31 ArgTrpThrLeuThraAlaProProGlyTyArg 41

Db 2686 TTGTGGACATGTGAAGTCCCAAAACAACATACCGC 2718

RESULT 12

US-09-925-302-168

Sequence 168, Application US/09925302

PATENT No. US20020044941A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antipodles

FILE REFERENCE: PA104

CURRENT APPLICATION NUMBER: US/09/925,302

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05918

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1998-03-12

NUMBER OF SEQ ID NOS: 896

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 168

LENGTH: 1148

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1076)

OTHER INFORMATION: n equals a,t,g, or c

US-09-925-302-168

Alignment Scores:

Pred. No.: 0.144 Length: 1148

Score: 81.00 Matches: 15

Percent Similarity: 58.82% Conservative: 5

Best Local Similarity: 44.12% Mismatches: 14

Query Match: 34.03% Indels: 0

DB: 10 Gaps: 0

US-09-874-198-1 (1-41) x US-09-925-302-168 (1-1148)

QY 8 ProgluProAlaPheGlyArgLeuAlaSerProGlyPheProGlyGluTyAlaAsnAsp 27

Db 139 CCTCAGAACTTATTGGGAGGTGACTTCCCTCTTCCCAAGCCTTACCCCAACAC 198

QY 28 GlnGluArgArgTrpThrLeuThraAlaProProGlyTyArg 41

Db 199 TTGGAACAACACACTGTGATCAGTCCCGGAGTACAGG 240

RESULT 13

US-09-919-497-39

Sequence 39, Application US/09919497

PATENT No. US2002010662A1

GENERAL INFORMATION:

APPLICANT: Mutter, George L.

TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

FILE REFERENCE: B0801/7225

CURRENT APPLICATION NUMBER: US/09/919,497

CURRENT FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: US 60/221,735

PRIOR FILING DATE: 2000-07-31

NUMBER OF SEQ ID NOS: 100

SOFTWARE: PatentIn version 3.0

SEQ ID NO 39

LENGTH: 1480

TYPE: DNA

ORGANISM: Homo sapiens

US-09-919-497-39

Alignment Scores:

Pred. No.: 0.194 Length: 1480

Score: 81.00 Matches: 17
Percent Similarity: 48.89% Conservative: 5
Best Local Similarity: 37.78% Mismatches: 13
Query Match: 34.03% Indels: 10
DB: 10 Gaps: 1

US-09-874-198-1 (1-41) x US-09-919-497-39 (1-1480)

QY 5 PolystyreneProgluProvalPhe-----glyarg 14
DB 142 CCCAACTACACAGACCCGTTCTTCGTCGAGGAGGATGTGAAGGGGGAATCAGGTTAC 201
QY 15 LeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTrrPhleu 34
DB 202 GTGGCAAGTGAAGGGGTTCCTCCCAACTCTTACCCCTTAATAGAGAGTCATCTGGACCATTA 261
QY 35 ThrAlaProProGly 39
DB 262 ACGGTCCCGAGGGC 276

14 880-107-2196
Sequence 2196, Application US/09880107
Patent No. US20020142981A1

GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2196
LENGTH: 1480
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 L33799
US-09-880-107-2196

Alignment Scores:
No.: 0.194 Length: 1480
Similarity: 48.89% Matches: 17
Best Local Similarity: 37.78% Conservative: 5
Query Match: 34.03% Mismatches: 13
DB: 10 Indels: 10
Gaps: 1

US-09-874-198-1 (1-41) x US-09-880-107-2196 (1-1480)

QY 5 PolystyreneProgluProvalPhe-----glyarg 14
DB 142 CCCAACTACACAGACCCGTTCTTCGTCGAGGAGGATGTGAAGGGGGAATCAGGTTAC 201
QY 15 LeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTrrPhleu 34
DB 202 GTGGCAAGTGAAGGGGTTCCTCCCAACTCTTACCCCTTAATAGAGAGTCATCTGGACCATTA 261
QY 35 ThrAlaProProGly 39
DB 262 ACGGTCCCGAGGGC 276

RESULT 15
US-09-808-602-92
Sequence 92, Application US/09808602
Patent No. US20020151151A1

GENERAL INFORMATION:

APPLICANT: Vernet, Corine A
APPLICANT: Fernandes, Elma
APPLICANT: Shinkels, Richard A
APPLICANT: Herrman, John L
APPLICANT: Majumder, Kumud
APPLICANT: Mishra, Vishnu
APPLICANT: Mezes, Peter S
APPLICANT: MacDougall, John
TITLE OF INVENTION: No. US20020151151A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 15966-697 CIP
CURRENT APPLICATION NUMBER: US/09/808,602
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 09/800,198
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: 60/186,596
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 92
LENGTH: 2386
TYPE: DNA
ORGANISM: Homo sapiens
US-09-808-602-92

Alignment Scores:
Pred. No.: 0.337 Length: 2386
Score: 81.00 Matches: 15
Percent Similarity: 58.82% Conservative: 5
Best Local Similarity: 44.12% Mismatches: 14
Query Match: 34.03% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x US-09-808-602-92 (1-2386)

QY 8 ProgluProvalPheGlyArgLeuAlaSerProGlyGluTyrAlaAsnAsp 27
DB 115 CCTGAGAGTATTTGGGAGGAGTGACTTCCCTCTGTCCTCCCAAGCTTACCCCAACAC 174
QY 28 GlnGluArgArgTrrPhleuThrAlaProProGlyTyrArg 41
DB 175 TTTGAACAACACAGTGTGATCACAGTCCCGACGGGATACAGG 216

RESULT 16
US-09-880-107-2256
Sequence 2256, Application US/09880107
Patent No. US20020142981A1

GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2256
LENGTH: 2493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 M14058
US-09-880-107-2256

Alignment Scores:
Pred. No.: 0.355 Length: 2493
Score: 81.00 Matches: 15

US-08-866-650-4

Alignment Scores:

Pred. No.:	65.8	Length:	3919
Score:	68.00	Matches:	11
Percent Similarity:	65.52%	Conservative:	8
Best Local Similarity:	37.93%	Mismatches:	10
Query Match:	28.57%	Indels:	0
DB:	2	Gaps:	0

US-09-874-198-1 (1-41) x US-08-866-650-4 (1-3919)

QY 13 GLYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32

DB 2994 GGCCCTCATCCACGAGTCCCACTGCGCAGACAGTACCCAGACAGCAAGAGATGACACTTG 3053

QY 33 ThrLeuThralaProProGlyTyrArg 41

DB 3054 GAATACAGCGCCACCTCCGCGCACCGA 3080

US-09-021-287-4

Sequence 4, Application US/09021287

Patent No. 5981717

GENERAL INFORMATION:

APPLICANT: Greenspan, Daniel S

APPLICANT: Takahara, Kazuhiko

APPLICANT: Hoffman, Guy G

TITLE OF INVENTION: Mammalian Tolloid-Like Protein

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/021,287

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/866,650

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Berson, Bennett J

REGISTRATION NUMBER: 37094

REFERENCE/DOCKET NUMBER: 960296,93839

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3919 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 648..3689

OTHER INFORMATION: /product= "human mTll protein"

US-09-021-287-4

Alignment Scores:

Pred. No.: 65.8 Length: 3919

Score: 68.00 Matches: 11

Percent Similarity: 65.52%

Best Local Similarity: 37.93%

Query Match: 28.57%

DB: 2 Gaps: 0

US-09-874-198-1 (1-41) x US-09-021-287-4 (1-3919)

QY 13 GLYArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32

DB 2994 GGCCCTCATCCACGAGTCCCACTGCGCAGACAGTACCCAGACAGCAAGAGATGACACTTG 3053

QY 33 ThrLeuThralaProProGlyTyrArg 41

DB 3054 GAATACAGCGCCACCTCCGCGCACCGA 3080

RESULT 18

US-09-240-473-4

Sequence 4, Application US/09240473

Patent No. 6297011

GENERAL INFORMATION:

APPLICANT: Greenspan, Daniel S

APPLICANT: Takahara, Kazuhiko

APPLICANT: Hoffman, Guy G

TITLE OF INVENTION: Mammalian Tolloid-Like Protein

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/240,473

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Berson, Bennett J

REGISTRATION NUMBER: 37094

REFERENCE/DOCKET NUMBER: 960296,93839

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3919 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 648..3689

OTHER INFORMATION: /product= "human mTll protein"

US-09-240-473-4

Alignment Scores:

Pred. No.: 65.8 Length: 3919

Score: 68.00 Matches: 11

Percent Similarity: 65.52%

Best Local Similarity: 37.93%

Query Match: 28.57%

DB: 2 Gaps: 0

US-09-874-198-1 (1-41) x US-09-240-473-4 (1-3919)

```

; TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,408
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/034,471
; FILING DATE: 02-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: ATG-50038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3690 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-991-408-3

Alignment Scores:
Pred. No.: 61.6 Length: 3690
Score: 68.00 Matches: 11
Percent Similarity: 65.52% Conservative: 8
Best Local Similarity: 37.93% Mismatches: 10
Query Match: 28.57% Indels: 0
DB: 3 Gaps: 0

US-09-874-198-1 (1-41) x US-08-991-408-3 (1-3690)
13 GtArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32
1143 GGCTCATCCACGCTCCCACTGCGCCAGACAGTACCCAGACAGGAGAAAGATGCACCTGG 1202
DB 1143 GGCTCATCCACGCTCCCACTGCGCCAGACAGTACCCAGACAGGAGAAAGATGCACCTGG 1202

OY 33 ThrLeuThrAlaProProGlyTyrArg 41
DB 1203 GAATCAGCGCCACCTCCTGCGCCACCGA 1229

RESULT 15
US-09-432-473-3
; Sequence 3, Application US/09432473
; Patent No. 6365715
; GENERAL INFORMATION:
; APPLICANT: ARLETH, ANTHONY J.
; APPLICANT: WILLETTTE, ROBERT N.
; APPLICANT: ELSHOURBAGY, NABIL A.
; APPLICANT: LI, XIAOTONG
; TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE PROTEIN
; FILE REFERENCE: ATG-50038-D1
; CURRENT APPLICATION NUMBER: US/09/432,473
; CURRENT FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: 08/991,408
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: 60/034,471
```

```

; EARLIER FILING DATE: 1997-01-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 3
; LENGTH: 3690
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; US-09-432-473-3

Alignment Scores:
Pred. No.: 61.6 Length: 3690
Score: 68.00 Matches: 11
Percent Similarity: 65.52% Conservative: 8
Best Local Similarity: 37.93% Mismatches: 10
Query Match: 28.57% Indels: 0
DB: 4 Gaps: 0

US-09-874-198-1 (1-41) x US-09-432-473-3 (1-3690)
OY 13 GtArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArgTyr 32
DB 1143 GGCTCATCCACGCTCCCACTGCGCCAGACAGTACCCAGACAGGAGAAAGATGCACCTGG 1202
DB 1203 GAATCAGCGCCACCTCCTGCGCCACCGA 1229

OY 33 ThrLeuThrAlaProProGlyTyrArg 41
DB 1203 GAATCAGCGCCACCTCCTGCGCCACCGA 1229

RESULT 16
US-08-866-650-4
; Sequence 4, Application US/08866650
; Patent No. 5939321
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Takahara, Kazuhiko
; APPLICANT: Hoffman, Guy G
; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Berson, Bennett J
; REGISTRATION NUMBER: 37094
; REFERENCE/DOCKET NUMBER: 960296, 93839
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 648..3689
; OTHER INFORMATION: /product= "human mt11 protein"
```


Query Match:	29.41%	Indels:	6
DB:	2	Gaps:	2
US-09-874-198-1 (1-41) x US-08-400-159-7 (1-4464)			
QY 2 ProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhePro 21			
Db 1650 CCATGGGGCCACACACGCCGGAG-----GCGTGTGTGCACAGCATCCCA 1606			
QY 22 GlyLeuTyrAlaAsnAspGlnGluArgGlyTrpThrLeuThrAlaProProGlyLys 40			
Db 1605 GGGCCCC--GCGTGTGACCCGACGCCATGCATCACTCTCAAGGCCCGCCAGGGCAC 1552			

```

US-08-611-729A-7/c
Sequence 7, Application US/08611729A
Patent No. 6004924
GENERAL INFORMATION:
APPLICANT: Ish-Horowitz, David
APPLICANT: Henrique, Domingos M.P.
APPLICANT: Lewis, Julian H.
APPLICANT: Myat, Anna M.
APPLICANT: Fleming, Robert J.
APPLICANT: Arlavanis-Tsakonas, Spyridon
APPLICANT: Mann, Robert S.
APPLICANT: Gray, Grace E.
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
NUMBER OF INVENTION: SRRATE GENE AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penille & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611.729A
FILING DATE: 06-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7526-037
ELECTRONIC COMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 332..4102
US-08-611-729A-7

Alignment Scores:
Pred. No.: 44.7 Length: 4483
Score: 70.00 Matches: 18
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 46.15% Mismatches: 12
Query/Match: 29,41% Gaps: 6
DB: 3
2

```

US-09-874-198-1 (1-41) x US-08-611-729A-7 (1-4483)

Oy	2	ProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhePro	21
Db	1550	CCATGGGGCCACACACCGCGAGG-----GCTGCTGGCAGCATCCCA	1606
Oy	22	GlycIuTyrAlaAsnAspGInGluArgArgTrpThrLeuThrAlaProGlyTyr	40
Db	1605	GGCCCC--GGCTGTGACCCGCGACCATCATCACTTCGACGGCCCCGGCAGGGAC	1552

```

RESULT 11
US-08-872-757-1
: Sequence 1, Application US/06872757
: Patent No. 6258584
: GENERAL INFORMATION:
: APPLICANT: PROCKOP, Darwin J.
: APPLICANT: Hojima, Yoshio
: APPLICANT: Li, Shi-Wu
: APPLICANT: Sieron, Aleksander
: TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
: TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/872,757
: FILING DATE: 10-JUN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/609,187
: FILING DATE: 01-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Halluin, Albert P.
: REGISTRATION NUMBER: 25,227
: REFERENCE/DOCKET NUMBER: 8389-028-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-854-3660
: TELEFAX: 415-854-3694
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2457 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2190
US-08-872-757-1
Alignment Scores:
Pred. NO.: 30.1 Length: 2457
Score: 69.00 Matches: 13
Percent Similarity: 58.62% Conservative: 4
Best Local Similarity: 44.83% Mismatches: 12
Query Match: 28.99% Indels: 0
DB: 4 Gaps: 0
US-09-874-198-1 (1-41) x US-08-872-757-1 (1-2457)
OY 13 GlyArgLeuAlaSerProGlyPheProGlyIleTyrAlaAsnAspGlnGluATGATGTTP 32
||||| ::: ||||| |::: ||||| |::: |||||

```


EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
ID NO: 37
LENGTH: 985
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (633)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (642)
OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-37
Alignment Scores:
Pred. No.: 1.68 Length: 985
Score: 76.00 Matches: 15
Percent Similarity: 62.07% Conservative: 3
Best Local Similarity: 51.72% Mismatches: 11
Query Match: 31.93% Indels: 0
DB: 4 Gaps: 0
US-09-874-198-1 (1-41) x US-09-227-357-37 (1-985)
QY 12 PhgGlaArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgArg 31
DB 690 TTGGGAACCTGAGAGCCCTGGATGCGCAGATACGACGACGACGATGCGATTCGCC 631
32 TrpThrLeuThrAlaProProGlyTyr 40
630 GTTACTCTCAGACCCCGCCAGACCAV 604
RESULT 6
US-08-839-008-4
Sequence 4, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1504 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-839-008-4
Alignment Scores:
Pred. No.: 4.58 Length: 1504
Score: 74.00 Matches: 16
Percent Similarity: 48.89% Conservative: 6
Best Local Similarity: 35.56% Mismatches: 13
Query Match: 31.09% Indels: 10
DB: 2 Gaps: 1
US-09-874-198-1 (1-41) x US-08-839-008-4 (1-1504)
QY 5 ProGlyTrpProGluProValPhe-----GlyArg 14
DB 122 CCCAACTACGAGACCTGTGTTCTGTGCGGAGGAGCGTACCGGAGTCAAGGTTAC 181
QY 15 LeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeu 34
DB 182 GTGGCAAGTAGGAGTTTCCCAACCTCTACCCCAACAGAGTGCATCTGACAAATT 241
QY 35 ThrAlaProProGly 39
DB 242 ACGGTGCCCGAGGGG 256
RESULT 7
US-08-839-008-6
Sequence 6, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk


```

: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,350B
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Wolfe, Susan A
: REGISTRATION NUMBER: 33,568
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ. ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4360 base pairs
: type: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: NT-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Rattus rattus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 94..3963
: US-08-470-350B-1
:
: Alignment Scores:
: Pred. No.: 1.17 length: 4360
: Score: 83.50 Matches: 16
: Percent Similarity: 41.18 Conservative: 5
: Best Local Similarity: 31.37% Mismatches: 17
: Query Match: 35.08% Indels: 13
: DB: 1 Gaps: 1
:
: US-09-874-198-1 (1-41) x US-08-470-350B-1 (1-4360)
:
: Oy 4 GlyProLysTrpProGluProValPhe----- 12
: Db 2566 GGTGCAATTGGACAAAGCTCTCTTGTGAACATACTTGTGGAGGTTTCTGACTGA 2625
: Oy 13 -----GlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnSpInGuLary 30
: Db 2626 CTCTCTGGGCAATTTCCTACCTACTACCTGAGGAGGAGCTATCTAATATGACGAGATGT 2685
: Oy 31 ArgTrpLeuThrAlaProProGlyTyrArg 41
: Db 2686 TTGTGGACATTGAGATCCCAACAACTACCGC 2718
:
: US-08-839-008-8
: Sequence 8, Application US/08839008
: Patent No. 5916758
: GENERAL INFORMATION:
: APPLICANT: Hurtle, Mark R
: APPLICANT: McDonnell, Peter C
: APPLICANT: McNulty, Dean E
: APPLICANT: Rosen, Craig A
: APPLICANT: Siemens, Ivo R
: APPLICANT: Young, Peter R
: APPLICANT: Yue, Tian-Li
: TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-Apr-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-Nov-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-839-008-8

Alignment Scores:
Pred. NO.: 0.691 Length: 1480
Score: 81.00 Matches: 17
Percent Similarity: 48.89% Conservative: 5
Best Local Similarity: 37.78% Mismatches: 13
Query Match: 34.03% Indels: 10
DB: 2 Gaps: 1

US-09-874-198-1 (1-41) x US-08-839-008-8 (1-1480)

QY 5 ProlystirProGluProValphe-----GlyArg 14
||| ::| ||||| ||| |||
Db 142 CCCAACTCACACCAACCCGCTGTCGCGGAGGATGGAAGGGGAAATCAGTTAC 201
15 leuAlasrProGlyPheProGlyGluTyraLasnaSPGInguarAgtrPrThLeu 34
::| ||||| ||||| ||||| ||| :||| ||||| ::|
Db 202 GTGGCAATGTGAGGGGTTCCCACTCCTACCCCTTATAAGAGTGCACTGTGACCATTA 261
QY 35 ThrAlaProProGly 39
||| ||| ||| |||
Db 262 ACGGTCCCCGAGGGC 276

RESULT 3
US-09-381-779-1
; Sequence 1, Application US/09381779
; Patent No. 6190868
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan M.
; APPLICANT: Deem, Michael W.
; APPLICANT: Simpson, John W.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING A NUCLEIC ACID SEQUENCE
; FILE REFERENCE: Cura-8 US 15966-508
; CURRENT APPLICATION NUMBER: US/09/381,779
; CURRENT FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 0.U.S.N. 60/054,887
; PRIOR FILING DATE: 1997-08-07
; PRIOR APPLICATION NUMBER: PCT/US98/16548
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2492
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-381-779-1

Alignment Scores:

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 11, 2003, 11:09:54 ; Search time 74 Seconds
(without alignments)
169.916 Million cell updates/sec

Title: US-09-874-198-1
Perfect score: 238
Sequence: 1 TPLGPKMPVPVFGRLASPGF.....GEVANDQERRWLTAPPGR 41

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Command line parameters:

-MODEL=frame+ p2n model -DRV=xld
-Q/cgcn2.1/USPTO.spool/US09874198/runat.10012003_092044_2595/app_query.fasta.1.199
-DB-Issued_Patents_NA -QFMT=fastap -SUFIX=p2n.rni -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=60 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=20
-MODE=LOCAL -OUPMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09874198.cgcn1.1.31.0runat.10012003_092044_2595 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

1: Issued_Patents_NA.*
2: /cgcn2.6/ptodata/1/ina/5A.COMB.seq.*
3: /cgcn2.6/ptodata/1/ina/5B.COMB.seq.*
4: /cgcn2.6/ptodata/1/ina/5A.COMB.seq.*
5: /cgcn2.6/ptodata/1/ina/5B.COMB.seq.*
6: /cgcn2.6/ptodata/1/ina/Backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	35.1	4360	1	US-08-470-350B-1
2	81	34.0	1480	1	US-08-839-008-8
3	81	34.0	2492	4	US-09-381-779-1
4	78	32.8	1537	2	US-08-839-008-1
5	76	31.9	985	4	US-09-227-357-37
6	74	31.1	1504	2	US-08-839-008-4
7	74	31.1	1506	2	US-08-839-008-6
8	70	29.4	3955	4	US-09-214-278-4
9	70	29.4	4464	2	US-08-400-159-7
10	70	29.4	4483	3	US-08-611-729A-7
11	69	29.0	2457	4	US-08-872-757-1
12	69	29.0	2487	1	US-08-377-292-1

13	68	28.6	3546	4	US-08-872-757-3	Sequence 3, Appl
14	68	28.6	3690	3	US-08-991-408-3	Sequence 3, Appl
15	68	28.6	3690	4	US-09-432-473-3	Sequence 3, Appl
16	68	28.6	3919	2	US-08-866-650-4	Sequence 4, Appl
17	68	28.6	3919	2	US-09-021-287-4	Sequence 4, Appl
18	68	28.6	3919	4	US-09-240-473-4	Sequence 4, Appl
19	68	28.6	5145	3	US-08-991-408-1	Sequence 1, Appl
20	68	28.6	5145	4	US-09-432-473-1	Sequence 1, Appl
21	67	28.2	4771	2	US-08-866-650-2	Sequence 2, Appl
22	67	28.2	4771	2	US-09-021-287-2	Sequence 2, Appl
23	67	28.2	4771	4	US-09-240-473-2	Sequence 2, Appl
24	65.5	27.5	1802	4	US-09-032-523-5	Sequence 5, Appl
25	64	26.9	3371	4	US-09-116-473-1	Sequence 1, Appl
26	63	26.5	1886	4	US-09-484-970B-127	Sequence 127, App
27	63	26.5	2730	3	US-08-936-135-17	Sequence 17, Appl
28	63	26.5	2766	3	US-08-936-135-3	Sequence 3, Appl
29	63	26.5	2772	3	US-08-936-135-1	Sequence 1, Appl
30	63	26.5	2781	3	US-08-936-135-19	Sequence 19, Appl
31	63	26.5	3471	4	US-09-116-473-3	Sequence 3, Appl
32	63	26.5	3652	3	US-08-936-135-5	Sequence 5, Appl
33	63	26.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
34	62	26.1	2089	4	US-09-079-431B-5	Sequence 5, Appl
35	62	26.1	2254	4	US-09-079-431B-3	Sequence 3, Appl
36	62	26.1	2257	4	US-09-079-431B-1	Sequence 1, Appl
37	61.5	25.8	1636	4	US-09-039-198A-1	Sequence 1, Appl
38	61.5	25.8	1636	4	US-08-877-599-1	Sequence 1, Appl
39	61.5	25.8	1636	4	US-09-267-574-1	Sequence 1, Appl
40	61.5	25.8	1643	2	US-08-486-839-3	Sequence 3, Appl
41	61.5	25.8	1643	3	US-09-151-011-3	Sequence 3, Appl
42	61.5	25.8	1643	4	US-09-343-623-3	Sequence 3, Appl
43	61.5	25.8	1656	4	US-09-039-198A-3	Sequence 3, Appl
44	61.5	25.8	1656	4	US-08-877-599-3	Sequence 3, Appl
45	61.5	25.8	1656	4	US-09-267-574-3	Sequence 3, Appl
46	61.5	25.8	1713	2	US-08-486-839-5	Sequence 5, Appl
47	61.5	25.8	1713	3	US-09-151-011-5	Sequence 5, Appl
48	61.5	25.8	1713	4	US-09-343-623-5	Sequence 5, Appl
49	61	25.6	530	3	US-08-758-662-4	Sequence 4, Appl
50	61	25.6	838	1	US-07-590-894C-1	Sequence 1, Appl
51	61	25.6	3529	3	US-08-936-135-7	Sequence 7, Appl
52	61	25.6	4524	2	US-08-845-998-7	Sequence 7, Appl
53	61	25.6	4524	4	US-09-206-537-7	Sequence 7, Appl
54	61	25.6	4524	3	US-09-430-854-7	Sequence 7, Appl
55	61	25.6	4718	3	US-08-936-135-9	Sequence 9, Appl
56	61	25.6	4733	3	US-08-936-135-11	Sequence 11, Appl
57	61	25.6	4765	3	US-08-936-135-21	Sequence 21, Appl
58	61	25.6	4769	3	US-08-936-135-13	Sequence 13, Appl
59	61	25.6	4780	3	US-08-936-135-23	Sequence 23, Appl
60	61	25.6	4784	3	US-08-936-135-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-470-350B-1
; Sequence 1, Application US/08470350B
; Patent No. 5684126
GENERAL INFORMATION:
APPLICANT: LI, Xiao
TITLE OF INVENTION: Snyder, Solomon H
TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 11, 2003, 11:09:24 ; Search time 2190 Seconds
(without alignments)
303.203 Million cell updates/sec

Title: US-09-874-198-1
Perfect score: 238
Sequence: 1 TPLGKWPPEVFGRLASPGF.....GEYANDQERRMTLTAPPGYR 41

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Matched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Command line parameters:
-MODEL=frim+ p2n.model -DEV=xlp
-O=gcgn2.1/uspro.spool/us09874198/runat.10012003_092044_2571/app_query.fasta.1.199
-DB=EST-QFMT-fastap -SUFRTX=p2n.rst -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsunb2 -TRANS=human40.cdi -DIST=60
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=20 -MODE=LOCAL
-OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09874198.gcgn2.1.1716@runat.10012003_092044_2571 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEDOUT=120
-HARN_TIMOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estlu: *
5: em_estlov: *
6: em_estlpl: *
7: em_estro: *
8: em_hic: *
9: gD_estl: *
10: gD_est2: *
11: gD_est3: *
12: gD_est4: *
13: gD_est5: *
14: gD_est6: *
15: em_estfun: *
16: em_estcom: *
17: gD_gss: *
18: em_gss_hum: *
19: em_gss_luv: *
20: em_gss_pin: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	796	14	B0650766 AGENCOURT
2	238	100.0	911	14	B0654231 AGENCOURT
3	238	100.0	913	14	B0648641 AGENCOURT
4	238	100.0	921	14	B0648719 AGENCOURT
5	238	100.0	939	14	B0644570 AGENCOURT
6	238	100.0	975	14	B0646367 AGENCOURT
7	238	100.0	1035	14	B0653181 AGENCOURT
8	183	76.9	348	10	B0656455 AGENCOURT
9	183	76.9	357	10	B0670802 AGENCOURT
10	183	76.9	416	14	B0671372 AGENCOURT
11	183	76.9	546	9	AA237253 AGENCOURT
12	183	76.9	552	9	AA238260 AGENCOURT
13	183	76.9	586	9	AA238260 AGENCOURT
14	183	76.9	600	12	BF532409 AGENCOURT
15	183	76.9	699	10	B0653589 AGENCOURT
16	183	76.9	729	9	AI226038 AGENCOURT
17	183	76.9	752	12	BF233497 AGENCOURT
18	183	76.9	843	13	BF236712 AGENCOURT
19	183	76.9	884	12	BF236712 AGENCOURT
20	183	76.9	884	12	BF236712 AGENCOURT
21	177	74.4	523	9	AI327510 AGENCOURT
22	173	72.7	450	9	AA208933 AGENCOURT
23	170	71.4	678	12	BF233993 AGENCOURT
24	153	64.3	487	9	AI048966 AGENCOURT
25	139	58.4	620	14	B0678079 AGENCOURT
26	125	52.5	598	9	AI391397 AGENCOURT
27	114	47.9	175	9	AA212846 AGENCOURT
28	93	39.1	224	12	B0663462 AGENCOURT
29	93	39.1	382	14	R11746 AGENCOURT
30	93	39.1	387	14	R11746 AGENCOURT
31	93	39.1	389	10	B0672700 AGENCOURT
32	93	39.1	390	10	B0672630 AGENCOURT
33	93	39.1	409	14	R12521 AGENCOURT
34	93	39.1	416	14	T97116 AGENCOURT
35	93	39.1	457	14	H15611 AGENCOURT
36	93	39.1	550	14	H19143 AGENCOURT
37	93	39.1	604	14	B0660916 AGENCOURT
38	93	39.1	619	9	AI099203 AGENCOURT
39	93	39.1	621	9	AL134380 AGENCOURT
40	93	39.1	630	10	B0621182 AGENCOURT
41	93	39.1	645	10	BE048313 AGENCOURT
42	93	39.1	646	10	AW957587 AGENCOURT
43	93	39.1	648	10	B0659136 AGENCOURT
44	93	39.1	653	10	AW238809 AGENCOURT
45	93	39.1	677	12	BF532377 AGENCOURT
46	93	39.1	698	10	B0623493 AGENCOURT
47	93	39.1	711	10	AW954607 AGENCOURT
48	93	39.1	719	10	B0640327 AGENCOURT
49	93	39.1	731	14	B0647443 AGENCOURT
50	93	39.1	734	14	B0771154 AGENCOURT
51	93	39.1	790	13	B1330963 AGENCOURT
52	93	39.1	894	13	B1913343 AGENCOURT
53	92	38.7	496	13	B1318546 AGENCOURT
54	91.5	38.4	530	13	B1775020 AGENCOURT
55	91.5	38.4	742	9	AI460182 AGENCOURT
56	90	37.8	503	14	B0739185 AGENCOURT
57	89	37.4	551	12	B0234368 AGENCOURT
58	89	37.4	744	10	AW211385 AGENCOURT
59	88.5	37.2	391	10	B0672811 AGENCOURT
60	87	36.6	360	14	D70256 CELK101H3F

ALIGNMENTS

RESULT 1
LOCUS B0650766 796 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8347134 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284012
5', mRNA sequence.
ACCESSION B0650766
VERSION B0650766
KEYWORDS EST.
SOURCE B0650766.1 GI:21774938
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2480 row: o column: 21
High quality sequence stop: 505.
Location/Qualifiers
1. 796
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6284012"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 142 a 264 c 243 g 144 t 3 others
ORIGIN

Alignment Scores:
Pied. No.: 6,39e-18 Length: 796
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x B0650766 (1-796)

QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 77 ACCCCCTGGGCGCCGAAGTGCTGAACTGTGTGGGGCGCTGGCAATCCCGGCTTT 136
|||||

QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40
|||||
Db 137 CCAGGGGAGATGCCAATGACGAGAGCGGCGCTGACCTGACACCCCGGCTTAC 196
|||||

QY 41 Arg 41
|||
Db 197 CGC 199

RESULT 2
LOCUS B0654231 911 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8493289 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6295026

ACCESSION B0654231
VERSION B0654231.1 GI:21778403
KEYWORDS EST.
SOURCE B0654231.1
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2511 row: a column: 11
High quality sequence start: 5
High quality sequence stop: 461.
Location/Qualifiers
1. 911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6295026"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 178 a 286 c 279 g 162 t 6 others
ORIGIN

Alignment Scores:
Pied. No.: 7,35e-18 Length: 911
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x B0654231 (1-911)

QY 1 ThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
|||||
Db 77 ACCCCCTGGGCGCCGAAGTGCTGAACTGTGTGGGGCGCTGGCAATCCCGGCTTT 136
|||||

QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40
|||||
Db 137 CCAGGGGAGATGCCAATGACGAGAGCGGCGCTGACCTGACACCCCGGCTTAC 196
|||||

QY 41 Arg 41
|||
Db 197 CGC 199

RESULT 3
LOCUS B0648641 913 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8299192 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271539
5', mRNA sequence.
ACCESSION B0648641
VERSION B0648641.1 GI:21772813

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Eutheria: Chordata: Cranialata: Vertebrata: Euteleostomi:
AUTHORS Eukaryota: Eutheria: Primates: Catarrhini: Homiidae: Homo.
TITLE 1. (bases 1 to 913)
JOURNAL NIH-MGC <http://mhc.nci.nih.gov/>.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2448 row: h column: 04
High quality sequence stop: 549.
Location/Qualifiers
1. 913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="IMAGE:6271539"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 193 a 298 c 255 g 166 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7.37e-18 Length: 913
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
-874-198-1 (1-41) x B0648641 (1-913)
QY 1 ThnProLenglyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 65 ACCCCCTTGGGCGGAGAGTGGCTGAACCTGTGTGGGGCGCTGGCATCCCCGGCTTT 124
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40
DB 125 CCAGGGGAGATGCAATGACAGAGGCGGCGCTGACCTGATGACCCCGGCTTAC 184
QY 41 Arg 41
DB 185 CGC 187
RESULT 4
LOCUS B0648719 921 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_818306 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6283375
5', mRNA sequence.
ACCESSION B0648719
VERSION B0648719.1 GI:21772891.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
AUTHORS 1 (bases 1 to 921)
TITLE NIH-MGC <http://mhc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2479 row: e column: 08
High quality sequence stop: 609.
Location/Qualifiers
1. 921
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="IMAGE:6283375"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 172 a 300 c 285 g 164 t
ORIGIN
Alignment Scores:
Pred. No.: 7.44e-18 Length: 921
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-09-874-198-1 (1-41) x B0648719 (1-921)
QY 1 ThnProLenglyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 65 ACCCCCTTGGGCGGAGAGTGGCTGAACCTGTGTGGGGCGCTGGCATCCCCGGCTTT 124
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40
DB 125 CCAGGGGAGATGCAATGACAGAGGCGGCGCTGACCTGATGACCCCGGCTTAC 184
QY 41 Arg 41
DB 185 CGC 187
RESULT 5
LOCUS B0644570 939 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8490607 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6293770
5', mRNA sequence.
ACCESSION B0644570
VERSION B0644570.1 GI:21768742
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC <http://mhc.nci.nih.gov/>.
AUTHORS 1 (bases 1 to 939)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2497 row: f column: 11
High quality sequence stop: 551.
Location/Qualifiers

FEATURES

source

1..939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 189 a 308 c 273 g 166 t 3 others
ORIGIN

Alignment Scores:

Pred. No.: 7.59e-18 Length: 939
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ644570 (1-939)

QY 1 ThProLeuGlyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

|||||
Db 65 ACCCCCTTGGGGCCGAAGTGGCTGAACCTGTTCGGGGCCTGCATCCCCGGCTTT 124

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40

|||||
125 CCAGGGGAGATGCCAATGACGAGCGGCGCTGCACCTGACCTGACCCCGGCTAC 184

41 Arg 41
|||

Db 185 CGC 187

RESULT 6 BQ646367 975 bp mRNA linear EST 15-JUL-2002

LOCUS BQ646367
DEFINITION AGENCOURT_8488545 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6296145

5', mRNA sequence.
ACCESSION BQ646367
VERSION BQ646367.1 GI:21770539

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 975)
NIH-MGC <http://mgc.nci.nih.gov/>.

Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2503 row: 1 column: 14
High quality sequence stop: 504.
Location/Qualifiers

FEATURES

source

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2503 row: 1 column: 14
High quality sequence stop: 504.
Location/Qualifiers

BASE COUNT

198 a 320 c 286 g 171 t

ORIGIN

Alignment Scores:

Pred. No.: 7.9e-18 Length: 975
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ646367 (1-975)

QY 1 ThProLeuGlyProLystrProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

|||||
Db 65 ACCCCCTTGGGGCCGAAGTGGCTGAACCTGTTCGGGGCCTGCATCCCCGGCTTT 124

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProProGlyTyr 40

|||||
Db 125 CCAGGGGAGATGCCAATGACGAGCGGCGCTGCACCTGACCTGACCCCGGCTAC 184

QY 41 Arg 41
|||

Db 185 CGC 187

RESULT 7 BQ653181 1035 bp mRNA linear EST 15-JUL-2002

LOCUS BQ653181
DEFINITION AGENCOURT_8303990 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6269223

5', mRNA sequence.
ACCESSION BQ653181
VERSION BQ653181.1 GI:21777353

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1035)
NIH-MGC <http://mgc.nci.nih.gov/>.

Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2503 row: 1 column: 14
High quality sequence stop: 504.
Location/Qualifiers

1..975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 198 a 320 c 286 g 171 t
ORIGIN

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LICM2442 row: 9 column: 16
High quality sequence stop: 510.

FEATURES

source

Location/Qualifiers
1..1035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6269223"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 224 a 341 c 298 g 172 t
ORIGIN

Alignment Scores:

Pred. No.: 8,41e-18 Length: 1035
Score: 238.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-874-198-1 (1-41) x BQ653181 (1-1035)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyValrGluAlaSerProGlyPhe 20
|||||
65 ACCCCCTTGGGCGGAGTGGCTGAGCTGTCTCGGGCGCTGGCATCCCCGGCTTT 124
QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40
|||||
Db 125 CCAGGGGAGATGCAATGACACAGAGCGCGCTGACCTGACGACCCCGGCTTAC 184
QY 41 Arg 41
|||
Db 185 CCG 187

LOC 8
425

AA896425 238 bp mRNA linear EST 06-APR-1998
v331e05.r1 StrataGene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1297088 5' similar to TR:000187 000187 MASP-2 PROTEIN. ;
mRNA sequence.

ACCESSION AA896425
VERSION AA896425.1 GI:3032818
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 238)
AUTHORS Mairra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheinberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WASHU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Mairra M/Mouse EST Project
WASHU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:678136
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 197.

FEATURES

source

Location/Qualifiers
1..238
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1297088"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."
BASE COUNT 44 a 69 c 67 g 58 t
ORIGIN

Alignment Scores:

Pred. No.: 4.26e-12 Length: 238
Score: 183.00 Matches: 32
Percent Similarity: 85.37% Conservative: 3
Best Local Similarity: 78.05% Mismatches: 6
Query Match: 76.89% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x AA896425 (1-238)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyValrGluAlaSerProGlyPhe 20
|||||
Db 116 ACACCTTGGGCTTCAAGTGGCTGAACTGTATTTCGGGCGCTGGTGTACCTGGCTTC 175
QY 21 ProGlyGluTrpAlaAsnAspGlnGluArgTrpThrLeuThrAlaProProGlyTyr 40
|||||
Db 176 CCAGAGAAATGATGCTGACATCAAGATGATCCCTGACACTGACGACCCCGGATAC 235
QY 41 Arg 41
|||
Db 236 CCG 238

RESULT 9 348 bp mRNA linear EST 27-NOV-2001
BB870802
LOCUS BB870802
DEFINITION BB870802 RIKEN full-length enriched, pooled tissues, intestinal mucosa, etc. Mus musculus cDNA clone G630025E18 5', mRNA sequence.
ACCESSION BB870802
VERSION BB870802.1 GI:17117012
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 348)
AUTHORS Akimura,T., Arkawa,T., Carlinici,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Saito,K., Shibata,K., Shinagawa,A., Shiraki,T., Soabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T., Watanishi,A., Yasunishi,A., Yamatsuta,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)

contact: Yoshitake Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suheiho-cho, Tsutsumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,T., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

SOURCE

Location/Qualifiers

1. 699
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C730007H2"
/clone_lib="RIKEN full-length enriched, adult male liver tumor"
/sex="male"
/tissue_type="liver tumor"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
GAGAGAGAGCGCGCGCAACGAGATTTTCTTTTCTTTTNN 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'
GAGAGAGATTCGAGATTAATTAATCCGCCGCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluscript KS(+) after bulk excision from Lambda FIC I. Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge."

BASE COUNT 146 a 209 c 182 g 162 t
ORIGIN

Alignment Scores:
Pred. No.: 1,32e-11 Length: 699
Score: 183.00 Matches: 32
Percent Similarity: 85.37% Conservative: 3
Best Local Similarity: 78.05% Mismatches: 6
Query Match: 76.89% Indels: 0
DB: 10 Gaps: 0

US-09-874-198-1 (1-41) x BB653589 (1-699)

QY 1 ThrProLeuClYProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
DB 65 ACACCTTCGGGTTCGAAGTGCCTGAACCTGATTCGGGGCTGGTCCCTGCCTTC 124
QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgGlyTrpThrLeuThrAlaProProGlyTyr 40
DB 125 CCAGAGAGATGCTGACCATCAAGATCGATCTCGACACTGACCCCTGGCTAC 184

QY 41 Arg 41
DB 185 CGC 187

RESULT 17
A1226038
LOCUS
DEFINITION
A1226038 729 bp mRNA linear EST 29-OCT-1998
uJ08e08.y1 Sugano mouse liver mla Mus musculus cDNA clone
A1226038 5' similar to TR:000187 000187 MASP-2 PROTEIN. ;
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
A1226038.1 GI:3809091
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 729)
Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

JOURNAL

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:975658

COMMENT

Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 510.

FEATURES

SOURCE

1. 729
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1891334"
/clone_lib="Sugano mouse liver mla"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGGGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor (TCCTGCGCTACGCG), digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGCTCTTAAGAAGTGGC and 3' end
primer CGACCTGACGCTGAGACACA."

BASE COUNT 151 a 221 c 187 g 167 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 1,37e-11 Length: 729
Score: 183.00 Matches: 32
Percent Similarity: 85.37% Conservative: 3
Best Local Similarity: 78.05% Mismatches: 6
Query Match: 76.89% Indels: 0
DB: 9 Gaps: 0

US-09-874-198-1 (1-41) x A1226038 (1-729)

QY 1 ThProleuglyProlystrProgluProvalphheglyArgleualaaserProglyPhe 20

Db 127 ACACCTTCTGGGTCACAAAGTGCGCTGATATCGGGCGGCGCTGCTGCCCTGCTTC 186

QY 21 ProglygluTyrrAlaasnaspgInguArgTrpThreuthralaProproglyTyr 40

Db 187 CCAGAGAAATGATCTGACATCAAGATCGATCTCGGACACTGATGACACCCCTGGCTAC 246

QY 41 Arg 41

Db 247 CGC 249

RESULT 18

LOCUS BF233497 752 bp mRNA linear EST 14-NOV-2000

DEFINITION 602024289F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4159415 5',

SYN BF233497

SOURCE BF233497.1 GI:11143969

ORGANISM EST.

REFERENCE house mouse.

AUTHORS Mus musculus

TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

JOURNAL Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

COMMENT NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9437 row: b column: 24

High quality sequence stop: 739.

Location/Qualifiers

1. 752

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4159415"

/lab_host="NCI_CGAP_L19"

/note="DH10B (T1 phage-resistant)"

Site_2: Salt: Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 180 a 218 c 186 g 168 t

ORIGIN

Alignment Scores:

Pred. No.: 1,42e-11

Score: 183.00

Percent Similarity: 85.37%

Best Local Similarity: 78.05%

Query Match: 76.89%

DB: 12 Gaps: 0

US-09-874-198-1 (1-41) x BF233497 (1-752)

QY 1 ThProleuglyProlystrProgluProvalphheglyArgleualaaserProglyPhe 20

Db 86 ACACCTTCTGGGTCACAAAGTGCGCTGATATCGGGCGGCGCTGCTGCCCTGCTTC 145

QY 21 ProglygluTyrrAlaasnaspgInguArgTrpThreuthralaProproglyTyr 40

Db 146 CCAGAGAAATGATCTGACATCAAGATCGATCTCGGACACTGATGACACCCCTGGCTAC 205

QY 41 Arg 41

Db 206 CGC 208

RESULT 19

LOCUS B1146955 843 bp mRNA linear EST 05-JUL-2001

DEFINITION 602911589F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5052574 5',

SYN B1146955

SOURCE B1146955.1 GI:14606956

ORGANISM EST.

REFERENCE house mouse.

AUTHORS Mus musculus

TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

JOURNAL Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

COMMENT NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1143 row: a column: 23

High quality sequence stop: 767.

Location/Qualifiers

1. 843

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5052574"

/lab_host="NCI_CGAP_L19"

/note="DH10B (T1 phage-resistant)"

Site_2: Salt: Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 186 a 247 c 221 g 189 t

ORIGIN

Alignment Scores:

Pred. No.: 1.6e-11

Score: 183.00

Percent Similarity: 85.37%

Best Local Similarity: 78.05%

Query Match: 76.89%

DB: 13 Gaps: 0

US-09-874-198-1 (1-41) x B1146955 (1-843)

QY 1 ThProleuglyProlystrProgluProvalphheglyArgleualaaserProglyPhe 20

Db 31 ACACCTTCTGGGTCACAAAGTGCGCTGATATCGGGCGGCGCTGCTGCCCTGCTTC 90

QY 21 ProglygluTyrrAlaasnaspgInguArgTrpThreuthralaProproglyTyr 40

Db 91 CCAGAGAAATGATCTGACATCAAGATCGATCTCGGACACTGATGACACCCCTGGCTAC 150

QY 41 Arg 41

Db 151 CGC 153

RESULT 20

LOCUS BF236712 884 bp mRNA linear EST 14-NOV-2000

DEFINITION 602028006F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4153448 5',

ACCESSION mRNA sequence.
 VERSION BF236712
 KEYWORDS BF236712.1 GI:1150598
 EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 884)
 NIH-MGC <http://mgc.ncl.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9447 row: k column: 01
 High quality sequence step: 615.
 Location/Qualifiers

FEATURES

source
 1..884
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4163448"
 /clone_1lb="NCI-CGAP-L19"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI-CGAP Library."
 BASE COUNT 191 a 256 c 252 g 185 t
 ORIGIN

Alignment Scores:

Pred. No.:	1.68e-11	Length:	884
Score:	183.00	Matches:	32
Percent Similarity:	85.378	Conservative:	3
Best Local Similarity:	78.058	Mismatches:	6
Query Match:	76.89%	Indels:	0
DB:	12	Gaps:	0

US-09-874-198-1 (1-41) x BF236712 (1-884)

```

1  TThProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
32  ACACCTCTGTGGTTCACAAAGCTGCTGAACCTGATTCGGGGCGCTGGTGCCTGCCTGCCTG 91

QY  21  ProGlyGluTrpAlaAsnAspGlnGluArgArgTrpThrLeuThrAlaProProGlyTyr 40
   ||| ::||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
Db   92  CCAGAGAAGATGCTGACATCAAGATCGATCCTGGACACTGACTGACACCCCTGGCTAC 151

QY  41  Arg 41
   |||
Db   152  CGC 154
```

Search completed: January 11, 2003, 12:46:28
 Job time : 2197 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 11, 2003, 11:06:19 ; Search time 351 Seconds

(without alignments)
263.054 Million cell updates/sec

Title: US-09-874-198-1

Perfect score: 238
Sequence: 1 TPLGKWEPEYVGRSLASPGF.....GEYANDQERRWTLTAPGVR 41

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 08
Maximum Match 100%

Listing first 60 summaries

Command line parameters:

-MODEL=frame+.p2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09874199/runat.10012003.092043.2552/app_query.fasta.1.199
-DB=N.Geneseq.101002 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=60 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptc -NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874198.ecgn.1.1.79.6runat.10012003.092043.2552 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGABEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	238	100.0	2472	24	AAD24224	Human MASP-2 CDNA
2	238	100.0	3106	23	AA855321	Human encoding novel
3	93.5	39.3	810	22	ABA06477	Human CDNA SEQ ID
4	93.5	39.3	810	22	AA531599	CDNA encoding novel
5	93.5	39.3	810	23	ABK43587	Human encoding novel
6	93	39.1	363	21	AA001735	Human secreted pro
7	93	39.1	397	21	AAC00227	Human secreted pro
8	93	39.1	363	22	AAH43512	CDNA encoding prot
9	93	39.1	385	22	AAE83985	Human serine prote
10	91.5	38.4	1988	22	AAE74442	Human PRO11 nucleo
11	91.5	38.4	2127	22	AAE74440	Human PRO9 nucleo
12	91.5	38.4	2127	22	AAE74441	Human PRO10 nucleo
13	91.5	38.4	2143	22	AAE74443	Human PRO12 nucleo
14	91.5	38.4	3663	22	AAE74438	Human PRO7 nucleo
15	91.5	38.4	3879	22	AAE74439	Human PRO8 nucleo
16	91.5	38.4	4034	24	ABK28635	Human CDNA encodin
17	91.5	38.4	4198	24	AAD25344	Human SE26 DNA. H
18	84	35.3	9455	23	ABL10283	Human CDNA encodin
19	84	35.3	13036	23	ABL10282	Drosophila melanog
20	83.5	35.1	4344	24	ABK63805	Rat sequence diffe
21	83.5	35.1	4360	18	AAT44068	Rat von Ebner's gl
22	81	34.0	1148	21	AAE18149	Lung cancer associ
23	81	34.0	1480	24	ABN95698	Gene #2196 used to
24	81	34.0	1480	24	ABK64507	Human benign prost
25	81	34.0	1480	24	ABK35519	Human endometrial
26	81	34.0	2492	20	AAK24297	Human complement c
27	81	34.0	2493	22	ABN83114	Complement compone
28	81	34.0	2493	24	ABN95758	Gene #2256 used to
29	79	33.2	382	21	AAA45208	Human secreted exp
30	79	33.2	2070	22	AA500132	Human CDNA encodin
31	79	33.2	2073	22	AA500128	Human CDNA encodin
32	79	33.2	2073	22	AA500136	Human CDNA encodin
33	79	33.2	2154	22	AA500116	Human CDNA encodin
34	79	33.2	2157	22	AA500112	Human CDNA encodin
35	79	33.2	2157	22	AA500120	Human CDNA encodin
36	79	33.2	3252	22	AA500133	Human CDNA encodin
37	79	33.2	3255	22	AA500129	Human CDNA encodin
38	79	33.2	3255	22	AA500137	Human CDNA encodin
39	79	33.2	3336	22	AA500117	Human CDNA encodin
40	79	33.2	3339	22	AA500113	Human CDNA encodin
41	79	33.2	3339	22	AA500121	Human CDNA encodin
42	79	33.2	3579	22	AA500138	Human CDNA encodin
43	79	33.2	3582	22	AA500127	Human CDNA encodin
44	79	33.2	3582	22	AA500135	Human CDNA encodin
45	79	33.2	3663	22	AA500115	Human CDNA encodin
46	79	33.2	3666	22	AA500111	Human CDNA encodin
47	79	33.2	3666	22	AA500119	Human CDNA encodin
48	79	33.2	3666	22	AA500130	Human CDNA encodin
49	79	33.2	3669	22	AA500126	Human CDNA encodin
50	79	33.2	3669	22	AA500134	Human CDNA encodin
51	79	33.2	3750	22	AA500114	Human CDNA encodin
52	79	33.2	3753	22	AA500110	Human CDNA encodin
53	79	33.2	3753	22	AA500118	Human CDNA encodin
54	79	33.2	4446	22	AA500131	Human CDNA encodin
55	78	32.8	1482	19	AAV22658	Nucleic acid encod
56	78	32.8	1537	18	AAT72652	Human smooth muscl
57	78	32.8	1537	20	AAK90357	Human smooth muscl
58	78	32.8	1542	24	ABK35135	Human CDNA encodin
59	78	32.8	1580	21	AAC78046	Human cancer assoc
60	78	32.8	1768	22	AAK94126	Human full-length

ALIGNMENTS

RESULT 1
AAD24224
ID AAD24224 standard; cDNA: 2472 BP.

CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 3106 BP; 848 A; 846 C; 798 G; 614 T; 0 other;

Alignment Scores:

Pred. No.: 1,36e-18 Length: 3106
Score: 238.00 Matches: 41
Local Similarity: 100.00% Conservatve: 0
Match: 100.00% Mismatches: 0
23 Indels: 0
Gaps: 0

US-09-874-198-1 (1-41) x AAS85321 (1-3106)

QY 1 ThrProLeuGlyProLysTrpProGluProValPheGlyArgLeuAlaSerProGlyPhe 20

DB 2255 ACCGCTTGAGCGCCGAGTGGCTGAACCTGTTCGGGCGCTGGCATCCCGGCTTT 2314

QY 21 ProGlyGluTyrAlaAsnAspGlnGluArgArgTyrThrLeuThrAlaProGlyTyr 40

DB 2315 CCAGGGGAGATGCCATGACACGAGAGCGGCGCTGACCTGACCTGACCCCGGCTAC 2374

QY 41 Arg 41

DB 2375 CGC 2377

RESULT 3

ABA06477/C
ID ABA06477 standard: cDNA; 810 BP.

XX ABA06477;

DT 10-JAN-2002 (first entry)

XX Human cDNA SEQ ID NO: 143.

DE Human; gene therapy; neural disorder; immune system disorder;
muscular disorder; reproductive disorder; gastrointestinal disorder;
pulmonary disorder; cardiovascular disorder; renal disorder;
proliferative disorder; inflammation; ss.

XX Homo sapiens.

OS WO200154474-A2.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01349.

XX 31-JAN-2000; 2000US-179065P.

XX 04-FEB-2000; 2000US-180628P.

XX 24-FEB-2000; 2000US-184664P.

XX 02-MAR-2000; 2000US-186350P.

XX 16-MAR-2000; 2000US-189874P.

XX 17-MAR-2000; 2000US-190076P.

XX 18-APR-2000; 2000US-198123P.

XX 19-MAY-2000; 2000US-205515P.

XX 07-JUN-2000; 2000US-209467P.

XX 28-JUN-2000; 2000US-214886P.

XX 30-JUN-2000; 2000US-215135P.

XX 07-JUL-2000; 2000US-216647P.

XX 07-JUL-2000; 2000US-216880P.

PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225265P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226688P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232387P.
PR 14-SEP-2000; 2000US-232388P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236337P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.

PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249287P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
PA (HUMAN-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-476161/51.
DR P-PSDB: ABB10255.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
XX condition -
XX
XX Claim 1. SEQ ID NO: 143; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
SQ Sequence 810 BP; 181 A; 189 C; 249 G; 187 T; 4 other;

Pred. No.: 0.0402
Score: 93.50
Percent Similarity: 46.154
Best Local Similarity: 40.388
Query Match: 39.298
DB: 22
US-09-874-198-1 (1-41) x ABA06477 (1-810)
QY 5 ProllystP-----ProgluProval----- 11
Db 606 CCCWCTGGGATTCAAGGAGCCCGTTCGCGCGAGTGATCCCAAT 547
QY 12 -----PheglYArgLeuAlaSerProGlyPheProGlytYlYAlaAsnAspGln 29
Db 546 GCCACCAACGCGCGCATCGCTCTCCAGGCTTCCGCGGACACACACAACTCACC 487
QY 30 ArgArgTrpThrLeuThrAlaProProGlyTyrArg 41
Db 486 TGTCACTGGCTGCTGAGGCTCTGAGGCGCACCG 451
RESULT 4
AAS31599/C
ID AAS31599 standard; cDNA; 810 BP.
XX
AC AAS31599;
XX
DT 04-DEC-2001 (first entry)
XX
DE cDNA encoding novel human calcium-binding protein #23.
XX
XX Human: calcium-binding protein; calcium flux; neurological disease;
KW immune dysfunction; digestive disorder; neoplastic disease;
KW blood disorder; infectious disease; gene therapy; immunosuppressive;
KW antitumor; cytostatic; vasotropic; antibacterial; mototropic;
KW virulence; ss.
XX
OS Homo sapiens.
XX
PN WO200155304-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01302.
FE
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR	14-AUG-2000	2000US-02257578
PR	14-AUG-2000	2000US-02257578
PR	18-AUG-2000	2000US-02262719
PR	22-AUG-2000	2000US-02266681
PR	22-AUG-2000	2000US-02266688
PR	22-AUG-2000	2000US-02271882
PR	23-AUG-2000	2000US-02277009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-02292827
PR	01-SEP-2000	2000US-02293443
PR	01-SEP-2000	2000US-02293444
PR	01-SEP-2000	2000US-02293445
PR	05-SEP-2000	2000US-02295509
PR	05-SEP-2000	2000US-02295513
PR	06-SEP-2000	2000US-02304357
PR	06-SEP-2000	2000US-02304368
PR	08-SEP-2000	2000US-02312442
PR	08-SEP-2000	2000US-02312443
PR	08-SEP-2000	2000US-02312444
PR	14-SEP-2000	2000US-02323387
PR	14-SEP-2000	2000US-02323398
PR	14-SEP-2000	2000US-02323399
PR	14-SEP-2000	2000US-02324000
PR	14-SEP-2000	2000US-02324401
PR	14-SEP-2000	2000US-02330663
PR	14-SEP-2000	2000US-02330664
PR	14-SEP-2000	2000US-02330665
PR	21-SEP-2000	2000US-02342223
PR	21-SEP-2000	2000US-02342224
PR	21-SEP-2000	2000US-0234223
PR	25-SEP-2000	2000US-02349977
PR	25-SEP-2000	2000US-02349978
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-02358834
PR	27-SEP-2000	2000US-02358836
PR	29-SEP-2000	2000US-02363377
PR	29-SEP-2000	2000US-02363637
PR	29-SEP-2000	2000US-02363638
PR	29-SEP-2000	2000US-02363639
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-02376802
PR	02-OCT-2000	2000US-02376937
PR	02-OCT-2000	2000US-02376938
PR	02-OCT-2000	2000US-02377039
PR	02-OCT-2000	2000US-02377040
PR	13-OCT-2000	2000US-02393515
PR	13-OCT-2000	2000US-02393517
PR	20-OCT-2000	2000US-02409660
PR	20-OCT-2000	2000US-02410221
PR	20-OCT-2000	2000US-02411785
PR	20-OCT-2000	2000US-02411787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-02418265
PR	01-NOV-2000	2000US-0244617
PR	01-NOV-2000	2000US-02464474
PR	08-NOV-2000	2000US-02464765
PR	08-NOV-2000	2000US-02464766
PR	08-NOV-2000	2000US-02465577
PR	08-NOV-2000	2000US-02465578
PR	08-NOV-2000	2000US-0246558
PR	08-NOV-2000	2000US-02466039
PR	08-NOV-2000	2000US-02466052

PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246611	
PR	08-NOV-2000	2000US-0246613	
PR	17-NOV-2000	2000US-0249207	
PR	17-NOV-2000	2000US-0249208	
PR	17-NOV-2000	2000US-0249209	
PR	17-NOV-2000	2000US-0249210	
PR	17-NOV-2000	2000US-0249211	
PR	17-NOV-2000	2000US-0249212	
PR	17-NOV-2000	2000US-0249213	
PR	17-NOV-2000	2000US-0249214	
PR	17-NOV-2000	2000US-0249215	
PR	17-NOV-2000	2000US-0249216	
PR	17-NOV-2000	2000US-0249217	
PR	17-NOV-2000	2000US-0249218	
PR	17-NOV-2000	2000US-0249219	
PR	17-NOV-2000	2000US-0249220	
PR	17-NOV-2000	2000US-0249221	
PR	17-NOV-2000	2000US-0249222	
PR	17-NOV-2000	2000US-0249223	
PR	17-NOV-2000	2000US-0249224	
PR	17-NOV-2000	2000US-0249225	
PR	17-NOV-2000	2000US-0249226	
PR	17-NOV-2000	2000US-0249227	
PR	17-NOV-2000	2000US-0249228	
PR	17-NOV-2000	2000US-0249229	
PR	17-NOV-2000	2000US-0249230	
PR	01-DEC-2000	2000US-0250160	
PR	01-DEC-2000	2000US-0250391	
PR	05-DEC-2000	2000US-0251030	
PR	05-DEC-2000	2000US-0251068	
PR	05-DEC-2000	2000US-0251988	
PR	05-DEC-2000	2000US-0256719	
PR	06-DEC-2000	2000US-0251479	
PR	08-DEC-2000	2000US-0251856	
PR	08-DEC-2000	2000US-0251868	
PR	08-DEC-2000	2000US-0251869	
PR	08-DEC-2000	2000US-0251869	
PR	08-DEC-2000	2000US-0251990	
PR	11-DEC-2000	2000US-0254097	
PR	05-JAN-2001	2001US-0259678	
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI: 2001-465568/50.		
XX			
DR	P-PSDB; AAU19914.		
XX			
PT	Isolated nucleic acid molecule encoding a calcium-binding protein is		
PT	used in preventing, treating or ameliorating a medical condition -		
XX			
PS	Claim 4; SEQ ID NO 33; 542pp; English.		
XX			
CC	The present invention relates to the isolation of novel human		
CC	calcium-binding proteins (AAU19892-AAU19969), and cDNA and genomic		
CC	sequences encoding for these proteins. The sequences of the invention		
CC	are useful in the diagnosis, prevention and/or prognosis of diseases		
CC	associated with aberrant calcium flux. Such disorders include		
CC	neurological diseases (e.g. amyotrophic lateral sclerosis, ALS),		
CC	immune dysfunction (e.g. severe combined immunodeficiency, SCID),		
CC	digestive disorders (e.g. irritable bowel syndrome, IBS), neoplastic		
CC	disease (e.g. cancer), blood disorders (e.g. haemophilia), and/or		
CC	infectious disease (e.g. acquired immunodeficiency syndrome, AIDS). The		
CC	novel calcium-binding proteins are also useful as screening tools to		
CC	identify antagonists and/or agonists that may enhance or inhibit		
CC	activities mediated by calcium-binding proteins. The polynucleotides of		
CC	the invention are also useful in gene therapy. AAS31577-AAS31654		
CC	represent cDNA sequences encoding for the novel human calcium-binding		
CC	proteins.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
XX			
SQ	Sequence 810 BP; 181 A; 189 C; 249 G; 187 T; 4 other;		
Alignment Scores:			
Pred. NO.:	0.0402	Length:	810
Score:	93.50	Matches:	21

Percent Similarity: 46.15%
 Best Local Similarity: 40.38%
 Query Match: 39.29%
 DB: 22
 Gaps: 2

US-09-874-198-1 (1-41) x AAS31599 (1-810)

QY 5 ProlytP-----ProgiuProval----- 11
 DB 606 CCCWTCGGGATTCGAAGGAGCCGCTGTCATGCTGCGGCGGAGATGATCCGCAAT 547
 QY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAspGlnGlu 29
 DB 546 GCCACACACCGCGCGATCTCTCCAGGCTCCCGGCAACTACAGACACATCCAC 487
 QY 30 ArgArgTyrPheLeuThrAlaProProGlyTyrArg 41
 DB 486 TGTCACTGGCTGCTGAGCTCTGAGGCGCACGCG 451
 AC ABK43587/c
 XX ABK43587 standard; cDNA: 810 BP.
 DT 05-JUN-2002 (first entry)
 XX
 DE DNA encoding novel central nervous system protein #167.
 XX
 OS Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
 XX hyperproliferative disorder; neoplasm; cardiovascular disorder;
 PN cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
 XX acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
 XX adenocarcinoma; reproductive system disorder; testicular feminisation;
 XX endocrine disorder; diabetes; cancer; leukemia; neovascularisation;
 XX respiratory disorder; renal disorder; kidney failure; blood disorder;
 XX myocardial infarction; wound healing; cell proliferation; skin aging;
 XX food additive; food preservative; gene therapy; gene; ss.
 OS Homo sapiens.
 XX
 PN WO20015318-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001: 2001MO-US01332.
 PR 31-JAN-2000: 2000US-0179065.
 PR 24-FEB-2000: 2000US-0180628.
 PR 24-FEB-2000: 2000US-0184664.
 PR 02-MAR-2000: 2000US-0186350.
 PR 16-MAR-2000: 2000US-0189874.
 PR 17-MAR-2000: 2000US-0190076.
 PR 18-APR-2000: 2000US-0198123.
 PR 19-MAY-2000: 2000US-0205515.
 PR 07-JUN-2000: 2000US-0209467.
 PR 28-JUN-2000: 2000US-0214886.
 PR 30-JUN-2000: 2000US-0215135.
 PR 07-JUL-2000: 2000US-0216647.
 PR 07-JUL-2000: 2000US-0216880.
 PR 11-JUL-2000: 2000US-0217487.
 PR 14-JUL-2000: 2000US-0217496.
 PR 26-JUL-2000: 2000US-0218290.
 PR 26-JUL-2000: 2000US-0220963.
 PR 26-JUL-2000: 2000US-0220964.
 PR 14-AUG-2000: 2000US-0224518.
 PR 14-AUG-2000: 2000US-0224519.
 PR 14-AUG-2000: 2000US-0225213.
 PR 14-AUG-2000: 2000US-0225214.
 PR 14-AUG-2000: 2000US-0225266.
 PR 14-AUG-2000: 2000US-0225267.
 PR 14-AUG-2000: 2000US-0225268.

PR 14-AUG-2000: 2000US-0225270.
 PR 14-AUG-2000: 2000US-0225447.
 PR 14-AUG-2000: 2000US-0225457.
 PR 14-AUG-2000: 2000US-0225758.
 PR 14-AUG-2000: 2000US-0225759.
 PR 18-AUG-2000: 2000US-0226279.
 PR 22-AUG-2000: 2000US-0226681.
 PR 22-AUG-2000: 2000US-0226688.
 PR 22-AUG-2000: 2000US-0227182.
 PR 23-AUG-2000: 2000US-0227209.
 PR 30-AUG-2000: 2000US-0228824.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229343.
 PR 01-SEP-2000: 2000US-0229344.
 PR 01-SEP-2000: 2000US-0229345.
 PR 05-SEP-2000: 2000US-0229509.
 PR 05-SEP-2000: 2000US-0229513.
 PR 06-SEP-2000: 2000US-0230437.
 PR 06-SEP-2000: 2000US-0230438.
 PR 08-SEP-2000: 2000US-0231242.
 PR 08-SEP-2000: 2000US-0231243.
 PR 08-SEP-2000: 2000US-0231244.
 PR 08-SEP-2000: 2000US-0231413.
 PR 08-SEP-2000: 2000US-0231414.
 PR 08-SEP-2000: 2000US-0232080.
 PR 08-SEP-2000: 2000US-0232081.
 PR 12-SEP-2000: 2000US-0231968.
 PR 14-SEP-2000: 2000US-0232397.
 PR 14-SEP-2000: 2000US-0232398.
 PR 14-SEP-2000: 2000US-0232399.
 PR 14-SEP-2000: 2000US-0232400.
 PR 14-SEP-2000: 2000US-0232401.
 PR 14-SEP-2000: 2000US-0233063.
 PR 14-SEP-2000: 2000US-0233064.
 PR 14-SEP-2000: 2000US-0233065.
 PR 21-SEP-2000: 2000US-0234223.
 PR 21-SEP-2000: 2000US-0234274.
 PR 25-SEP-2000: 2000US-0234997.
 PR 25-SEP-2000: 2000US-0234998.
 PR 26-SEP-2000: 2000US-0235484.
 PR 27-SEP-2000: 2000US-0235834.
 PR 27-SEP-2000: 2000US-0235836.
 PR 29-SEP-2000: 2000US-0236327.
 PR 29-SEP-2000: 2000US-0236367.
 PR 29-SEP-2000: 2000US-0236368.
 PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239935.
 PR 13-OCT-2000: 2000US-0239937.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241221.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.
 PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI: 2001-581633/65.

DR P-PSDB: AAU87257.

XX New isolated nucleic acid encoding a protein for diagnosing,
preventing, treating or ameliorating medical conditions and used as
food additives or preservatives -

PS Claim 1; SEQ ID NO 177; 837pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX novel central nervous system protein. (I) and polypeptides (II) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
CC angiodystrophic, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infection. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.:	0.0402	Length:	810
Score:	93.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	39.29%	Indels:	15
DB:	23	Gaps:	2

US-09-874-198-1 (1-41) x ABK43587 (1-810)

QY 5 ProlystTrp-----ProgluProval----- 11

DB 606 CCCWCTGGGATTCGAAGAGGCCGTTMTGCATCGCTTCGGCGGAGATCCCAAT 547

QY 12 -----PhedGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29

DB 546 GCCACACGCGCGCCGATCGTCTCCAGGCTCCCGGCACTACAGCAACACCTCACC 487

QY 30 ArgArgTyrPheThrAlaProProGlyTyrArg 41

DB 486 TGTCACTGCTGCTTGAGGCTCTTGAGGCCACGCG 451

RESULT 6

AAAC01735
ID AAC01735 standard; cDNA; 363 BP.

AC AAC01735;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 1733.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PR (GEST) GENSET.

PA Dumas Mline Edwards J, Duclert A, Giordano J;

PI WPI: 2000-500381/45.

DR P-PSDB: AAC01729.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 1733; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain

CC The sequences given in AAAH3512-22 encode novel human proteases PRS-1
CC to -11. These proteases can be administered in pharmaceutical compounds
CC for the treatment of diseases or conditions associated with decreased
CC expression of functional protease. Disorders which can be treated
CC include gastrointestinal (peptic esophagitis, indigestion and
CC gastritis), cardiovascular (hypertensive heart disease, myocardial
CC infarction), autoimmune/inflammatory (rheumatoid arthritis), cell
CC proliferative (arteriosclerosis, cancers), epithelial (eczema),
CC neurological (Huntington's disease), and reproductive (endometriosis)
CC disorders. The protein encoded by this sequence shows homology
CC with the human precursor of P100 serine protease of Ra-reactive factor.
XX
SQ Sequence 3863 BP; 923 A; 1063 C; 991 G; 886 T; 0 other:

Alignment Scores:

Pred. No.: 0.264 Length: 3863
Score: 93.00 Matches: 14
Percent Similarity: 67.74% Conservative: 7
Local Similarity: 45.16% Mismatches: 10
Match: 39.08% Indels: 0
Gaps: 22

US-09-874-198-1 (1-41) x AAAH3512 (1-3863)

QY 11 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArg 30
:::|||||::: |||||||::: ||| :||| |||
DB 164 ATGTTGGCCAGATCCAGTCGCTGCTATCCAGACTCCATCCAGTATTCAGAGGTG 223
CC

QY 31 ArgTfThleuThralaProProGlyTyrArg 41
||| :||| ||| |||:::||||
DB 224 ACTTGGAAATATCACTGTCCACAGATGGGTTTCGG 256
XX

RESULT 9

AAF83985

ID AAF83985 standard; DNA; 3895 BP.

AC AAF83985;

DT 06-AUG-2001 (first entry)

XX Human serine protease MASP-3 encoding DNA.
XX
DE Mannan-binding lectin associated serine protease-3; MASP-3; MBL; human;
XX mannan-binding lectin complement; infection; cancer; cytostatic;
KW immunomodulator; neuroprotective; anticonvulsive; antipneumatic;
KW antiatheritic; antianemic; antiinflammatory; antiporiatic; antidiabetic;
XX antiatherosclerotic; ds.

Homo sapiens.

Key Location/Qualifiers
FT 91..2277
FT /tag=a
FT /product="MASP-3"

WO200140451-A2.

07-JUN-2001.

30-NOV-2000; 2000WO-DK00659.

02-DEC-1999; 99DK-0001721.

21-JUL-2000; 2000DK-0001126.

(JENS/) JENSENIUS J C.

(THIE/) THIEL S.

Jensenius JC, Thiel S;

WPI: 2001-374820/39.

P-PSDB; AAB85060.

Novel pure mannan-binding lectin associated serine protease

PT polypeptides and polynucleotides encoding the protein, used for
PT treating reoxygenated ischemic tissues, mannon-binding lectin
PT deficiency, multiple sclerosis
XX
XX Claim 18; Page 88-96; 99pp: English.

XX The invention relates to a mannan-binding lectin (MBL) associated serine
CC protease-3 (MASP-3) polypeptide. MASP-3 exerts an inhibitory effect on
CC the complement activation particularly when bound to MBL/MASP-2 complexes
CC and directly activates complement system through binding to MBL. The
CC MASP-3 polypeptides, polynucleotides and modulators are useful for
CC preparing a pharmaceutical composition for treating aberrant MASP-3
CC activity such as infections, cancer, MBL-deficiency, disorders of the
CC immune system and reproductive system. The MASP-3 polypeptide is also
CC used for treating diseases associated with human immunodeficiency virus,
CC multiple sclerosis, myasthenia gravis, epilepsy, rheumatoid arthritis,
CC vasculitis, autoimmune hemolytic anemia, Crohn's disease, asthma,
CC diabetes, psoriasis, multiple myeloma, atherosclerosis etc. It is also
CC useful for inhibiting activation of C4 complement by inhibiting the MBL
CC pathway, inhibiting MASP-2 activity, inhibiting or treating an
CC inflammatory condition related to complement activation through MBL/MASP
CC complexes. MASP-3 is also useful for treating an inflammatory condition
CC resulting from an autoimmune condition after acute myocardial infarction
CC or brain ischemia. It is also useful for treating an individual suffering
CC from a disorder resulting from an imbalanced cytokine network. The
XX present sequence represents a DNA encoding a human MASP-3 polypeptide.

SQ Sequence 3895 BP; 952 A; 1068 C; 987 G; 888 T; 0 other:

Alignment Scores:

Pred. No.: 0.267 Length: 3895
Score: 93.00 Matches: 14
Percent Similarity: 67.74% Conservative: 7
Best Local Similarity: 45.16% Mismatches: 10
Query Match: 39.08% Indels: 0
DB: 22 Gaps: 0

US-09-874-198-1 (1-41) x AAF83985 (1-3895)

QY 11 ValPheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGluArg 30
:::|||||::: |||||||::: ||| :||| |||
DB 169 ATGTTGGCCAGATCCAGTCGCTGCTATCCAGACTCCATCCAGTATTCAGAGGTG 228
CC

QY 31 ArgTfThleuThralaProProGlyTyrArg 41
||| :||| ||| |||:::||||
DB 229 ACTTGGAAATATCACTGTCCACAGATGGGTTTCGG 261
XX

RESULT 10

AAF74442

ID AAF74442 standard; cDNA; 1988 BP.

AC AAF74442;

DT 09-MAY-2001 (first entry)

DE Human PRO11 nucleotide sequence SEQ ID NO:21.

XX Human; PRO; cytosolic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.

Homo sapiens.

WO200110902-A2.

15-FEB-2001.

11-AUG-2000; 2000WO-US21857.

11-AUG-1999; 99US-0148433.

10-AUG-2000; 2000US-0148433.

(CURA-) CURAGEN CORP.

```
XX Shinkets RA, Fernandes E;
PI
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70541.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 38-40; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 1988 BP; 419 A; 695 C; 512 G; 362 T; 0 other:

Alignment Scores:
Pred. No.: 0.189 Length: 1988
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
DB: 22 Gaps: 2

US-09-874-198-1 (1-41) x AAF74442 (1-1988)

OY 5 ProlystRTP-----ProGluProVal----- 11
DB 1384 CCTTCTGGGATTCAAGAGCCGCTGCTGCTTCGCGCGGAGTGATCCGCAAT 1443
OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GGCACCAACCGCGCGCTGCTCTCCAGGCTCCCGGCACTACAGCAACAACCTCACC 1503
OY 30 ArgArgTTrpThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCCACGCG 1539
AAAF74440
ID AAF74440 standard; cDNA; 2127 BP.
XX
XX AAF74440;
AC
XX
XX 09-MAY-2001 (first entry)
DT
XX
XX Human PRO9 nucleotide sequence SEQ ID NO:17.
DE
XX
XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.
XX
OS Homo sapiens.
XX
XX WO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX 11-AUG-2000; 2000MO-US21857.
XX
XX 11-AUG-1999; 99US-0148433.
PR
```

```
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Fernandes E;
PI
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70539.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 33-35; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 2127 BP; 430 A; 764 C; 557 G; 375 T; 1 other:

Alignment Scores:
Pred. No.: 0.204 Length: 2127
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
DB: 22 Gaps: 2

US-09-874-198-1 (1-41) x AAF74440 (1-2127)

OY 5 ProlystRTP-----ProGluProVal----- 11
DB 1384 CCTTCTGGGATTCAAGAGCCGCTGCTGCTTCGCGCGGAGTGATCCGCAAT 1443
OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GGCACCAACCGCGCGCTGCTCTCCAGGCTCCCGGCACTACAGCAACAACCTCACC 1503
OY 30 ArgArgTTrpThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAGGCTCCTCGAGGCCACGCG 1539
RESULT 12
AAAF74441
ID AAF74441 standard; cDNA; 2127 BP.
XX
XX AAF74441;
AC
XX
XX 09-MAY-2001 (first entry)
DT
XX
XX Human PRO10 nucleotide sequence SEQ ID NO:19.
DE
XX
XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;
KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.
XX
OS Homo sapiens.
XX
XX WO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX
```

PF 11-AUG-2000: 2000MO-US21857.
XX
XX 11-AUG-1999: 99US-0148433.
PR 10-AUG-2000: 2000US-0148433.
XX
PA (CURA-) CURAGEN CORP.
PI Shimkets RA, Fernandes E;
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70540.
XX
PT Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 35-37; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
XX polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
XX X is an integer from 1 to 17). PROX polypeptides have cytostatic,
XX immunomodulatory and reproduction activities, and can be used in gene
XX therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
XX nucleic acids and antibodies are useful in the manufacture of a
XX medicament for treating a syndrome associated with a PROX-associated
XX disorder, e.g. a cell proliferation and/or differentiation disorder
XX (e.g. cancer or immune associated disorders) and a gestational disease
XX (e.g. pre-clampsia). They are also used for screening for a modulator of
XX activity or of latency or predisposition to a PROX-associated disorder.
XX C AAF74432 to AAF74448 encode the specifically claimed human PROX
XX polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 2127 BP; 431 A; 758 C; 557 G; 380 T; 1 other;
XX
XX Alignment Scores:
Pred. No.: 0.204 Length: 2127
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
Gaps: 2
XX
US-09-874-198-1 (1-41) x AAF74441 (1-2127)
XX
OY 5 ProlystRTP-----ProgluProval----- 11
DB 1384 CCCTTCGGGATTCAAAGAGAGCCCGCTGCATCGCTGCGGCGAGTGATCCGCAAT 1443
12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLutYrAlaAsnAspGlnGlu 29
1444 GGCACACACCGCGCGATCGTCTCCAGGCTTCCCGGCAACTACACCAACCTCACC 1503
OY 30 ArgArgTrpThrLeuThrAlaProProGlyTYrArg 41
DB 1504 TGTCACTGGCTGCTTGAAGCTCCTGAGGCGCACGCG 1539
XX
XX RESULT 13
XX AAF74443
XX ID AAF74443 standard; cDNA; 2143 BP.
XX
XX AAF74443;
XX
XX 09-MAY-2001 (first entry)
XX
XX Human PRO12 nucleotide sequence SEQ ID NO:23.
XX
XX Human: PRO; PROX; cytostatic; immunomodulatory; reproduction;
XX gene therapy; cell proliferation; differentiation disorder; cancer;
XX immune associated disorder; gestational disease; pre-clampsia; ss.
XX
XX Homo sapiens.
XX
XX WO200110902-A2.
XX
XX PN

XX
XX 15-FEB-2001.
PD
XX
XX 11-AUG-2000: 2000MO-US21857.
PF
XX
XX 11-AUG-1999: 99US-0148433.
PR 10-AUG-2000: 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Fernandes E;
PI
XX
XX WPI: 2001-147509/15.
DR P-PSDB: AAB70542.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
XX
PS Claim 8: Page 41-43; 166pp; English.
XX
XX The present invention describes isolated nucleic acids encoding secreted
XX polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
XX X is an integer from 1 to 17). PROX polypeptides have cytostatic,
XX immunomodulatory and reproduction activities, and can be used in gene
XX therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
XX nucleic acids and antibodies are useful in the manufacture of a
XX medicament for treating a syndrome associated with a PROX-associated
XX disorder, e.g. a cell proliferation and/or differentiation disorder
XX (e.g. cancer or immune associated disorders) and a gestational disease
XX (e.g. pre-clampsia). They are also used for screening for a modulator of
XX activity or of latency or predisposition to a PROX-associated disorder.
XX C AAF74432 to AAF74448 encode the specifically claimed human PROX
XX polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX
SQ Sequence 2143 BP; 431 A; 769 C; 559 G; 383 T; 1 other;
XX
XX Alignment Scores:
Pred. No.: 0.206 Length: 2143
Score: 91.50 Matches: 21
Percent Similarity: 46.15% Conservative: 3
Best Local Similarity: 40.38% Mismatches: 13
Query Match: 38.45% Indels: 15
Gaps: 2
XX
US-09-874-198-1 (1-41) x AAF74443 (1-2143)
XX
OY 5 ProlystRTP-----ProgluProval----- 11
DB 1384 CCCTTCGGGATTCAAAGAGAGCCCGCTGCATCGCTGCGGCGAGTGATCCGCAAT 1443
12 -----PheGlyArgLeuAlaSerProGlyPheProGlyLutYrAlaAsnAspGlnGlu 29
1444 GGCACACACCGCGCGATCGTCTCCAGGCTTCCCGGCAACTACACCAACCTCACC 1503
OY 30 ArgArgTrpThrLeuThrAlaProProGlyTYrArg 41
DB 1504 TGTCACTGGCTGCTTGAAGCTCCTGAGGCGCACGCG 1539
XX
XX RESULT 14
XX AAF74438
XX ID AAF74438 standard; cDNA; 3863 BP.
XX
XX AAF74438;
XX
XX 09-MAY-2001 (first entry)
XX
XX Human PRO7 nucleotide sequence SEQ ID NO:13.
XX
XX Human: PRO; PROX; cytostatic; immunomodulatory; reproduction;
XX gene therapy; cell proliferation; differentiation disorder; cancer;
XX immune associated disorder; gestational disease; pre-clampsia; ss.
XX
XX
XX

OS Homo sapiens.
XX MO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX 11-AUG-2000; 2000MO-US21857.
XX
XX 11-AUG-1999; 99US-0148433.
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
PA
PI Shimkets RA, Fernandes E;
PI
DR WPI: 2001-147509/15.
DR P-PSDB; AAB70537.
XX
XX Nucleic acids encoding secreted polypeptides, designated PROX
XX polypeptides, useful for treating a syndrome associated with a
XX PROX-associated disorder, e.g. cancer -

Claim 8; Page 24-27; 16pp; English.

XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX

Sequence 3863 BP; 781 A; 1306 C; 1028 G; 747 T; 1 other;

Alignment Scores:

Pred. No.:	0.398	Length:	3863
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	38.45%	Indels:	15
DB:	22	Gaps:	2

-874-198-1 (1-41) x AAF74438 (1-3863)

5 ProLysTrp-----ProGluProVal----- 11
DB 1384 CCATCTGGGATTCMAAGGAGCCGTRATGCATGCTTGGCGGAGTGATCCGCAT 1443
QY 12 -----PhGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GCCACACCGCGCGCATGCTCTCCAGGCTTCCCGGCAACTACAGCAACACCTCACC 1503
QY 30 ArgArgTyrPheThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAAGCTCTCTGAGGGCCACGCG 1539

RESULT 15

AAE74439
ID AAE74439 standard; cDNA; 3879 BP.

XX AAE74439;

XX 09-MAY-2001 (first entry)

XX Human PRO8 nucleotide sequence SEQ ID NO:15.

XX Human; PRO; PROX; cytostatic; immunomodulatory; reproduction;

KW gene therapy; cell proliferation; differentiation disorder; cancer;
KW immune associated disorder; gestational disease; pre-clampsia; ss.

XX Homo sapiens.
XX
XX MO200110902-A2.
XX
XX 15-FEB-2001.
XX
XX 11-AUG-2000; 2000MO-US21857.
XX
XX 11-AUG-1999; 99US-0148433.
PR 10-AUG-2000; 2000US-0148433.
XX
XX (CURA-) CURAGEN CORP.
PA
PI Shimkets RA, Fernandes E;
PI
DR WPI: 2001-147509/15.
DR P-PSDB; AAB70538.

XX Nucleic acids encoding secreted polypeptides, designated PROX
PT polypeptides, useful for treating a syndrome associated with a
PT PROX-associated disorder, e.g. cancer -
PS
PS Claim 8; Page 28-31; 16pp; English.

XX The present invention describes isolated nucleic acids encoding secreted
CC polypeptides, designated PROX polypeptides (i.e. a PRO polypeptide where
CC X is an integer from 1 to 17). PROX polypeptides have cytostatic,
CC immunomodulatory and reproduction activities, and can be used in gene
CC therapy, and as PROX antagonists and PROX agonists. PROX polypeptides,
CC nucleic acids and antibodies are useful in the manufacture of a
CC medicament for treating a syndrome associated with a PROX-associated
CC disorder, e.g. a cell proliferation and/or differentiation disorder
CC (e.g. cancer or immune associated disorders) and a gestational disease
CC (e.g. pre-clampsia). They are also used for screening for a modulator of
CC activity or of latency or predisposition to a PROX-associated disorder.
CC AAF74432 to AAF74448 encode the specifically claimed human PROX
CC polypeptides PRO1 to PRO17 given in AAB70531 to AAB70547.
XX

Sequence 3879 BP; 782 A; 1311 C; 1030 G; 755 T; 1 other;

Alignment Scores:

Pred. No.:	0.4	Length:	3879
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	3
Best Local Similarity:	40.38%	Mismatches:	13
Query Match:	38.45%	Indels:	15
DB:	22	Gaps:	2

US-09-874-198-1 (1-41) x AAF74439 (1-3879)

5 ProLysTrp-----ProGluProVal----- 11
DB 1384 CCATCTGGGATTCMAAGGAGCCGTRATGCATGCTTGGCGGAGTGATCCGCAT 1443
QY 12 -----PhGlyArgLeuAlaSerProGlyPheProGlyGluTyrAlaAsnAspGlnGlu 29
DB 1444 GCCACACCGCGCGCATGCTCTCCAGGCTTCCCGGCAACTACAGCAACACCTCACC 1503
QY 30 ArgArgTyrPheThrLeuThrAlaProProGlyTyrArg 41
DB 1504 TGTCACTGGCTGCTTGAAGCTCTCTGAGGGCCACGCG 1539

RESULT 16

ABK28635
ID ABK28635 standard; cDNA; 4034 BP.

XX ABK28635;

XX 09-APR-2002 (first entry)

XX

DE Human cDNA encoding secreted protein SECP2.
 XX
 XX Human; ss: gene: SECP: antiinflammatory; cytostatic; cardiant;
 KM immunosuppressive; antiviral; anti-HIV; antiallergic; antineumatic;
 KM muscular active general; anticonvulsant; nootropic; neuroprotective;
 KM antiallergic; hypotensive; cardiovascular disorder; atherosclerosis;
 KM hypertension; myocardial infarction; autoimmune disorder;
 KM inflammatory disorder; AIDS; acquired immunodeficiency syndrome;
 KM allergy; rheumatoid arthritis; cell proliferative disorder; cancer;
 KM developmental disorder; Duchenne muscular dystrophy;
 KM neurological disorder; epilepsy; Alzheimer's disease.
 OS Homo sapiens.
 XX
 XX MO200198353-A2.
 PN
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US19862.
 XX
 XX 20-JUN-2000; 2000US-212890P.
 XX 23-JUN-2000; 2000US-213466P.
 PR 27-JUN-2000; 2000US-214601P.
 PR 31-JUL-2000; 2000US-222372P.
 PR 08-SEP-2000; 2000US-231435P.
 PR 15-SEP-2000; 2000US-232889P.
 XX
 XX (INCYTE GENOMICS INC.
 PA Hillman JL, Tang YT, Yue H, Elliott VS, Tribouley CM, Lee EA;
 XX Ramkumar J, Lal P, Xu Y, Warren BA, Hafalia AJA, Baughn MR;
 PI Azimzal Y, Batera S, Burford N, Yao MG, Nguyen DB, Lu DAM;
 PI Wella NK, Gandhi AR, Au-Young J, Patterson C;
 XX
 XX WPI: 2002-090431/12.
 DR P-PSDB: AA081976.
 XX
 XX Forty four human secreted proteins (referred to as SECP-1 to SECP-44),
 PT useful in the diagnosis, treatment and prevention of cardiovascular
 PT (e.g. atherosclerosis), autoimmune/inflammatory (e.g. allergies) and
 PT cell proliferative disorders -
 XX
 XX Claim 5; Page 166-167; 195pp; English.
 CC
 CC The invention relates to forty four human secreted proteins (referred to
 CC as SECP-1 to SECP-44) and the nucleic acids encoding them. Also
 CC included are a host cell transformed with the nucleic acid, a
 CC transgenic animal comprising the nucleic acid, an anti-SECP
 CC antibody, use of the SECP proteins in isolating agonists and antagonists
 CC of SECP activity and a method of isolating compounds which alter the
 CC expression of the SECP nucleic acid. The SECP polynucleotides and
 CC polypeptides are useful in the diagnosis, treatment and prevention of
 CC cardiovascular (e.g. atherosclerosis, hypertension, myocardial
 CC infarction), autoimmune/inflammatory (e.g. acquired immunodeficiency
 CC syndrome (AIDS), allergies, rheumatoid arthritis), cell proliferative
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular
 CC dystrophy), and neurological (e.g. epilepsy, Alzheimer's disease)
 CC disorders. Numerous other examples of each disorder are given in the
 CC specification. The present sequence is a cDNA encoding a SECP protein.
 CC
 XX Sequence 4034 BP; 840 A; 1305 C; 1062 G; 827 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 0.417 Length: 4034
 Score: 91.50 Matches: 21
 Percent Similarity: 46.15% Conservative: 3
 Best Local Similarity: 40.38% Mismatches: 13
 Query Match: 38.45% Indels: 15
 DB: 24 Gaps: 2
 US-09-874-198-1 (1-41) x ABK28635 (1-4034)
 OY 5 ProlystPr-----ProgiuProval----- 11

Db 1217 CCCTCTCGGATTCAAAGAGAGCCGCTGCTGCGGCGGAGATCCGCAAT 1276
 OY 12 -----PheGlyArgLeuAlaSerProGlyPheProGlyGlyTyrAlaAsnAspGlnGlu 29
 Db 1277 GCCACGACCGCCGCGATCGCTCTCCAGGCTCCGCGGCAACTACAGCAACACTCCACC 1336
 OY 30 ArgArgTrpPheThrLeuPheAlaProProGlyTyrArg 41
 Db 1337 TGTACACTGGCTGCTGAGGCTCTGAGAGCGCAGCGG 1372
 RESULT 17
 AAD25344
 ID AAD25344 standard; DNA; 4198 BP.
 AC AAD25344;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX
 XX Human SEZ6 DNA.
 DE
 XX Human; SEZ6; neural regeneration; seizure; infertility; gene therapy;
 KM stroke; Alzheimer's disease; Huntington's disease; myasthenia gravis;
 KM neurotrophic; neuroprotective; anticonvulsant; cerebroprotective; stroke;
 KM vasotropic; neuronal growth; growth factor-mediated chemotaxis; trauma;
 KM neurological disorder; muscular dystrophy; muscle injury; vulnery;
 KM amyotrophic lateral sclerosis; multiple sclerosis; ischaemia; diabetes;
 KM epilepsy; Parkinson's disease; sexual development; gene mapping;
 KM impotence; libido; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH CDS 191..2732
 FT /*tag= a
 FT /product= "Human SEZ6 protein"
 FT sig_peptide 191..262
 FT /*tag= b
 FT mat_peptide 263..2749
 FT /*tag= c
 FT /product= "Mature hSEZ6 protein"
 FT misc_feature 191..2749
 FT /*tag= d
 FT /note= "this region is specifically claimed as SEQ
 FT ID NO: 2 in claim 13 of the specification"
 XX
 XX WO200183552-A2.
 PN
 XX 08-NOV-2001.
 PD
 XX 17-APR-2001; 2001WO-US10809.
 PF
 XX 28-APR-2000; 2000US-200200P.
 PR
 XX (ELIL) LILLY & CO ELI.
 PA
 XX Su EW;
 PI
 XX WPI: 2002-082840/11.
 DR P-PSDB: AAE15853, AAE15854.
 XX
 XX Novel human SEZ6 polypeptide useful for inducing neural regeneration,
 PT inhibiting neural degeneration, preventing seizures, and for treating
 PT infertility, Alzheimer's disease, stroke, seizures, Huntington's
 PT disease -
 XX
 Claim 13; Page 117-118; 127pp; English.
 CC The invention relates to an isolated human SEZ6 (hSEZ6) polypeptide and
 CC its corresponding DNA molecule. SEZ6 is useful for enhancing neuronal
 CC growth, neurite outgrowth, neuronal regeneration, neuronal survival,
 CC growth factor-mediated chemotaxis, altered behavioural patterns, e.g.,
 CC sleep or eating disorders and for treating neurological disorders such

CC as trigeminal neuralgia, Bell's palsy, myasthenia gravis, muscular
CC dystrophy, muscle injury, invertebrate disk syndrome, thoracic outlet
CC destruction syndrome, amiotropic lateral sclerosis, multiple sclerosis,
CC ischemia associated with stroke, neuropathy associated with diabetes,
CC spinal cord trauma, facial nerve crush and other trauma, Huntington's
CC disease. SEZ6 is useful for treating a patient suffering from a
CC neurological disorder such as epilepsy, Alzheimer's disease, Parkinson's
CC disease, seizure related disorder or a disorder associated with stroke.
CC SEZ6 DNA is useful as probes for gene mapping and for detecting
CC transcription, translation and/or expression of hSEZ6 polypeptide in
CC human tissue. A transgenic animal is useful as an animal model in
CC research and drug development procedures, and for testing compounds or
CC other treatment modalities which may prevent, suppress or cure a
CC pathology or disease associated with hSEZ6 activities. SEZ6 is also
CC useful for treating abnormal primary or secondary sexual development,
CC e.g., impotence, infertility or reduced libido. The hSEZ6 polynucleotide
CC is useful for treating the above mentioned disorders by gene therapy
CC techniques. The present sequence is human SEZ6 DNA.

CC Sequence 4198 BP; 854 A; 1386 C; 1119 G; 839 T; 0 other;

CC Alignment Scores:

Seq. No.:	0.436	Length:	4198
Score:	91.50	Matches:	21
Percent Similarity:	46.15%	Conservative:	13
Best Local Similarity:	40.38%	Mismatches:	15
Query Match:	38.45%	Indels:	2
DB:	24	Gaps:	2

US-09-874-198-1 (1-41) x AAD25344 (1-4198)

QY 5 ProlystRP-----ProGIUProVal-----11

Db 1397 CCCTTCGGGATTCGAAGGAGCCGCTCGATCGCTCGCGCGGAGATCCGCAAT 1456

QY 12 -----PhcGLyAlgluAlaSerProGLyPheProGLyAlaAsnAspGLnGlu 29

Db 1457 GCCACGACCGCGCGCATCGTCTCCAGGCTCCGCGGCACTACACACACCTCACC 1516

QY 30 ArgArgTfPThrLeuThraIaProProGLyTYrArg 41

Db 1517 TGTCACTGGCTGCTGCTGAGGCTCCTGAGGCGCACGGG 1552

RESULT 18

ID ABL10283 standard; cDNA; 9455 BP.

XX ABL10283;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25331.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66180.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1; SEQ ID NO 25331; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA

CC sequences (AB101840-AB116175) and the encoded proteins

CC (AB101840-AB116175).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at http://wipo.int/pub/published_pcl_sequences.

XX SQ Sequence 9455 BP; 2333 A; 2397 C; 2484 G; 2241 T; 0 other;

QY 13 G1ArgLeuAlaSerProGLyPheProGLyAlaAsnAspGLnGluArgTfRP 32

Db 1738 GCGGTGCTGGATCTCCGGCTACTCTGCTCAGCGCGGACCGAAGTCACTCTCCCTGG 1797

QY 33 ThrLeuThraIaProProGLyTYrArg 41

Db 1798 CACTGACGCGCGCGTTTGGCTACCGG 1824

RESULT 19

ID ABL10282 standard; cDNA; 13036 BP.

XX ABL10282;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25328.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66179.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1: SEQ ID NO 25328; 21np + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB161757-AB161757).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 13036 BP; 3489 A; 3058 C; 3137 G; 3352 T; 0 other:

Alignment Scores:
NO.: 12 Length: 13036
Matches: 84.00
Conservative: 68.97%
Local Similarity: 55.17%
Query Match: 35.29%
DB: 23 Gaps: 0

US-09-874-198-1 (1-41) x ABL10282 (1-13036)
QY 13 GYArgLeuAlaSerProGlyPheProGlyGlyTyrAlaAsnAspGlnGluArgTyr 32
DB 3161 GGGCTGCTGGAGATCTCCGGCTATCTCGTACGCGGACCGAGAACGACTGCTGCTGG 3220
QY 33 ThrLeuThrAlaProProGlyTyrArg 41
DB 3221 CAACGTGACGGCGCGCTTGGCTACCG 3247

RESULT 20
ABK63805
ID ABK63805 standard; cDNA; 4344 BP.
XX
AC ABK63805;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #1712.
XX
KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
differential expression; centrilobular necrosis; steatosis.

Rattus norvegicus.

WO200210453-A2.
XX
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-US23872.
XX
PR 31-JUL-2000; 2000US-222040P.
XX
PR 02-NOV-2000; 2000US-244880P.
XX
PR 11-MAY-2001; 2001US-290029P.
XX
PR 15-MAY-2001; 2001US-290645P.
XX
PR 22-MAY-2001; 2001US-292336P.
XX
PR 06-JUN-2001; 2001US-295798P.
XX
PR 13-JUN-2001; 2001US-297457P.
XX
PR 19-JUN-2001; 2001US-298884P.
XX
PR 09-JUL-2001; 2001US-303459P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendlick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX
XX WPI; 2002-241625/29.
XX
XX Predicting toxic effects of compounds or the progression of these toxic
PT

PT effects by determining the changes in gene expression in tissues or
PT cells exposed to the toxin and comparing these to gene expression in
PT unexposed tissues or cells -
XX
PS Claim 1: Seq ID No 1712; 239pp; English.
XX
XX
CC The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic
CC effect of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression.
CC The method can also be used to identify an agent which modulates the
CC toxic response and predict cellular pathways that a compound modulates
CC in a cell. The method utilizes a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridizes to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity
CC is characterized by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent.
XX
SQ Sequence 4344 BP; 1035 A; 1099 C; 1103 G; 1107 T; 0 other:

Alignment Scores:
Pred. No.: 4.02 Length: 4344
Score: 83.50
Percent Similarity: 41.18% Matches: 16
Best Local Similarity: 31.37% Mismatches: 5
Query Match: 35.08% Indels: 13
DB: 24 Gaps: 1

US-09-874-198-1 (1-41) x ABK63805 (1-4344)
QY 4 GYProlysrProGlyuProValPhe----- 12
DB 2566 GGTCCAGTTTGGACAACTGCTCTTGTAAACTATCTTGAGAGTTTCTGACTGCA 2625
QY 13 -----GlyArgLeuAlaSerProGlyPheProGlyGlyTyrAlaAsnAspGlnGluArg 30
DB 2626 CTCTCTGGCAATTTTCTACCCATCTACTACCTGGAGCTATCTATAATGCCAGATGT 2685
QY 31 ArgTrrThrLeuThrAlaProProGlyTyrArg 41
DB 2686 TTGTGCAACATTGAGTCCCAACAACACTACCGC 2718

Search completed: January 11, 2003, 11:16:24
Job time : 358 secs

THIS PAGE BLANK (USPTO)